Package 'FlexDir'

January 20, 2025

Version 1.0

Title Tools to Work with the Flexible Dirichlet Distribution

Author Sonia Migliorati [aut],
Agnese Maria Di Brisco [aut, cre],
Matteo Vestrucci [aut]

Maintainer Agnese Maria Di Brisco <agnese.dibrisco@unimib.it>

Description Provides tools to work with the Flexible Dirichlet
distribution. The main features are an E-M algorithm for computing the maximum
likelihood estimate of the parameter vector and a function based on conditional
bootstrap to estimate its asymptotic variance-covariance matrix. It contains
also functions to plot graphs, to generate random observations and to handle
compositional data.

Depends R (>= 3.0.0)

Imports stats, graphics, utils, grDevices

License GPL (>= 2)

LazyData true

NeedsCompilation no

RoxygenNote 6.0.0

Suggests testthat

Repository CRAN

Date/Publication 2017-03-16 13:35:03

Contents

FD.aicbic   *Information Criterions of a Flexible Dirichlet Model*

Description

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) of a fitted Flexible Dirichlet model. An Information Criterion for one fitted model object for which a log-likelihood value can be obtained is defined as $-2 * log - likelihood + k * npar$, where $npar$ represents the number of parameters in the fitted model, and $k = 2$ for AIC, or $k = log(n)$ for BIC (n being the number of observations).

Usage

```
FD.aicbic(x)
```

Arguments

x   an object of class FDfitted, usually the result of `FD.estimation`.

See Also

`FD.estimation`, `FD.stddev`, `FD.barycenters`

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
FD.aicbic(results)
```

FD.amalgamation *Amalgamation*

Description

Given a matrix or a numeric dataframe, this function returns a composition where a set of specified columns is amalgamated together. The compositional operation of amalgamation provides sums of composition elements aimed at grouping homogeneous parts of the whole.

Usage

```
FD.amalgamation(data, columns, name = NULL)
```

Arguments

data   a matrix or a dataframe containing only variables to be transformed into compositional variables, after amalgamation.

columns   numeric vector containing the position of the columns to be amalgamated together.

name   string containing the name of the new column resulted from the amalgamation.

Details

Values must be positive. In case one row-entry (or more) is NA, the whole row will be returned as NA.

See Also

[FD.generate](), [FD.subcomposition](), [FD.normalization]()

Examples

```
data(oliveoil)
dataoil <- oliveoil
head(dataoil)
data <- FD.normalization(dataoil[,3:10])
head(data)
data.sub <- FD.subcomposition(data,c(1,3,4,5))
head(data.sub)
data.amalg <- FD.amalgamation(data,c(2,6,7,8),name='others')
head(data.amalg)
```

FD.barycenters   *Cluster Barycenters of a Flexible Dirichlet model*

Description

Cluster barycenters of a fitted Flexible Dirichlet distribution.

Usage

```
FD.barycenters(x)
```

Arguments

x   an object of class FDfitted, usually the result of `FD.estimation`.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

`FD.estimation`, `FD.clusterdistances`, `FD.moments`

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
FD.barycenters(results)
```

FD.clusterdistances   *Flexible Dirichlet Cluster Distances*

Description

Returns a measure of symmetrized Kullback-Leibler distance between mixture component densities of a fitted Flexible Dirichlet distribution.

Usage

```
FD.clusterdistances(x)
```

Arguments

x   an object of class FDfitted, usually the result of `FD.estimation`.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

`FD.estimation`, `FD.barycenters`, `FD.moments`

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
FD.clusterdistances(results)
```

FD.density   *The Flexible Dirichlet Density Function*

Description

Density function on the simplex for the Flexible Dirichlet distribution with parameters a, p and t.

Usage

```
FD.density(x, a, p, t)
```

Arguments

x	vector of a point on the simplex. It must sum to one.
a	vector of the non-negative alpha parameters.
p	vector of the clusters' probabilities. It must sum to one.
t	non-negative scalar tau parameter.

Details

Vectors x, a and p must be of the same length.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

`FD.theorcontours`, `FD.generate`

Examples

```
x <- c(0.1,0.25,0.65)
alpha <- c(12,7,15)
prob <- c(0.3,0.4,0.3)
tau <- 8
FD.density(x,alpha,prob,tau)
```

FD.estimation   *Flexible Dirichlet Estimation*

Description

Estimates the vector of parameters of a Flexible Dirichlet distribution through an EM-based maximum likelihood approach.

Usage

```
FD.estimation(data, normalize = F, iter.initial.SEM = 50,
  iter.final.EM = 100, verbose = T)
```

Arguments

data	a matrix or a dataframe containing only the variables in the model. Rows must sum to one, or `normalize` must be set TRUE.
normalize	if TRUE, each row of `data` will be divided by its own total to become a point of the simplex. Values in `data` must be positive.
iter.initial.SEM	
	number of iterations for the initial SEM step. Default to 50.
iter.final.EM	number of iterations for the final EM step. Default to 100.
verbose	if TRUE, the progression of the elaboration and the results will be printed on screen.

Details

The procedure is made up of four stages:

1. Clustering: The algorithm applies many different clustering rules to the dataset, in order to exploit the specific cluster patterns that the parameter structure of the model involves.

2. Labelling: Once the initial partitions are obtained, group labeling needs to be established because any clustering algorithm assigns the group labels randomly, but the FD cluster structure entails a precise labelling scheme.

3. Initial SEM: A Stochastic E-M algorithm is applied at every initial partition and every possible label permutation identified.

4. Final E-M: The previous step must be seen as a multiple initialization strategy. At this point only the best one is selected and a final E-M algorithm is used to find the point that maximizes the likelihood of the parameter vector.

Value

an object of class FDfitted. It's a list composed by:

alpha Estimated values of the parameter vector Alpha

p Estimated values of the parameter vector P

tau Estimated value of the parameter Tau

logL LogLikelihood

data Normalized dataset

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

[FD.generate](), [FD.stddev](), [FD.aicbic](), [FD.barycenters](), [FD.ternaryplot](), [FD.rightplot](), [FD.marginalplot]()

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
summary(results)
```

FD.generate   *The Flexible Dirichlet Random Generation*

Description

Random generation from the Flexible Dirichlet distribution with parameters a, p and t.

Usage

```
FD.generate(n, a, p, t)
```

Arguments

n	number of points on the simplex to be generated.
a	vector of the non-negative alpha parameters.
p	vector of the clusters' probabilities. It must sum to one.
t	non-negative scalar tau parameter.

Details

Vectors a and p must be of the same length. The Flexible Dirichlet distribution derives from the normalization of a basis of positive dependent random variables obtained by starting from a basis of independent equally scaled gamma random variables, and randomly allocating to the i-th element a further independent gamma random variable.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, 1–21.

See Also

FD.estimation, FD.density, FD.theorcontours, FD.subcomposition, FD.amalgamation

Examples

```
n <- 100
alpha <- c(12,7,15)
prob <- c(0.3,0.4,0.3)
tau <- 8
data <- FD.generate(n,alpha,prob,tau)
data
```

FD.marginalplot   *Marginal Plot of a Flexible Dirichlet*

Description

Histogram of the observed marginal variable and estimated density function of the marginal variable of a fitted Flexible Dirichlet distribution.

Usage

```
FD.marginalplot(x, var, zoomed = T, showgrid = T, showdata = T)
```

Arguments

x	an object of class FDfitted, usually the result of FD.estimation.
var	position of the variable to be plotted.
zoomed	if TRUE, shows only the area where most of the density is concentrated. If FALSE, shows the whole range [0;1].
showgrid	if TRUE, shows the axis and the labels. If FALSE, only the graph is printed.
showdata	if TRUE, prints the histogram of the data. If FALSE, shows only the density function.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

FD.estimation, FD.ternaryplot, FD.rightplot

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
FD.marginalplot(results, var=2)
FD.marginalplot(results, var=2, zoomed=FALSE, showgrid=TRUE, showdata=FALSE)
```

FD.moments *Flexible Dirichlet Moments*

Description

Moments of a fitted Flexible Dirichlet distribution. The function returns the mean and variance vectors and the covariance and correlation matrices.

Usage

```
FD.moments(x)
```

Arguments

x   an object of class FDfitted, usually the result of `FD.estimation`.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

`FD.estimation`, `FD.barycenters`, `FD.clusterdistances`

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
FD.moments(results)
```

```
FD.normalization          Normalization
```

Description

Given a matrix or a numeric dataframe, this function returns a composition (i.e. data summing up to 1).

Usage

```
FD.normalization(data)
```

Arguments

data   a matrix or a dataframe containing only variables to be transformed into compositional variables.

Details

Values must be positive. In case one row-entry (or more) is NA, the whole row will be returned as NA.

See Also

`FD.generate`, `FD.subcomposition`, `FD.amalgamation`

Examples

```
data(oliveoil)
dataoil <- oliveoil
head(dataoil)
data <- FD.normalization(dataoil[,3:10])
head(data)
data.sub <- FD.subcomposition(data,c(1,3,4,5))
head(data.sub)
data.amalg <- FD.amalgamation(data,c(2,6,7,8),name='others')
head(data.amalg)
```

FD.rightplot	*Right Triangle Plot of a Flexible Dirichlet*

Description

Right triangle plot and contour lines of the density function of a fitted Flexible Dirichlet distribution.

Usage

```
FD.rightplot(x, var = c(1, 2), zoomed = T, showgrid = T, showdata = T,
  nlevels = 10)
```

Arguments

x	an object of class FDfitted, usually the result of `FD.estimation`.
var	numeric vector containing the two variables to be plotted on the axis.
zoomed	if TRUE, shows only the area where most of the density is concentrated. If FALSE, shows the whole area of the right triangle.
showgrid	if TRUE, shows the axis and the labels. If FALSE, only the graph is printed.
showdata	if TRUE, prints the data points. If FALSE, shows only the contour lines.
nlevels	approximate number of contour lines to be drawn.

Details

The number of variables in the fitted model must be 3 to draw a plot on the right triangle.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

`FD.estimation`, `FD.ternaryplot`, `FD.marginalplot`

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
FD.rightplot(results)
FD.rightplot(results, var=c(3,2), zoomed=FALSE, showgrid=TRUE, showdata=FALSE, nlevels=3)
```

FD.stddev	*Standard Deviation of the ML estimators of a Flexible Dirichlet*

Description

Conditional Bootstrap evaluation of the standard errors of the maximum likelihood parameter estimates of a Flexible Dirichlet distribution.

Usage

```
FD.stddev(x, iter.bootstrap = 500)
```

Arguments

x   an object of class FDfitted, usually the result of `FD.estimation`.

`iter.bootstrap` number of iterations of the Bootstrap.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

`FD.estimation`, `FD.aicbic`, `FD.barycenters`

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
FD.stddev(results)
```

`FD.subcomposition`   *Subcomposition*

Description

Given a matrix or a numeric dataframe, this function returns a subcomposition made up of the specified columns.

Usage

```
FD.subcomposition(data, columns)
```

Arguments

data   a matrix or a dataframe containing only variables in the model.

columns   numeric vector containing the position of the columns to keep in the new composition.

Details

Values must be positive. In case one row-entry (or more) is NA, the whole row will be returned as NA.

See Also

`FD.generate`, `FD.amalgamation`, `FD.normalization`

Examples

```
data(oliveoil)
dataoil <- oliveoil
head(dataoil)
data <- FD.normalization(dataoil[,3:10])
head(data)
data.sub <- FD.subcomposition(data,c(1,3,4,5))
head(data.sub)
data.amalg <- FD.amalgamation(data,c(2,6,7,8),name='others')
head(data.amalg)
```

FD.ternaryplot   *Ternary Plot of a Flexible Dirichlet*

Description

Ternary plot and contour lines of the density function of a fitted Flexible Dirichlet distribution.

Usage

```
FD.ternaryplot(x, zoomed = T, showgrid = T, showdata = T, nlevels = 10)
```

Arguments

x	an object of class FDfitted, usually the result of `FD.estimation`.
zoomed	if TRUE, shows only the area where most of the density is concentrated. If FALSE, shows the whole area of the ternary diagram.
showgrid	if TRUE, shows the axis and the labels. If FALSE, only the graph is printed.
showdata	if TRUE, prints the data points. If FALSE, shows only the contour lines.
nlevels	approximate number of contour lines to be drawn.

Details

The number of variables in the fitted model must be 3 to draw a ternary plot.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

`FD.estimation`, `FD.rightplot`, `FD.marginalplot`

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
FD.ternaryplot(results)
FD.ternaryplot(results, zoomed=FALSE, showgrid=TRUE, showdata=FALSE, nlevels=3)
```

FD.theorcontours   *Contour Lines of a Flexible Dirichlet*

Description

Contour lines of a Flexible Dirichlet with given parameters on the ternary diagram or on the right triangle.

Usage

```
FD.theorcontours(a, p, t, type = "ternary", var = c(1, 2), zoomed = T,
  showgrid = T, nlevels = 10)
```

Arguments

a	vector of the non-negative alpha parameters.
p	vector of the clusters' probabilities. It must sum to one.
t	non-negative scalar tau parameter.
type	string indicating whether to plot the contour lines on a ternary diagram `'ternary'`, or on a right triangle plot `'right'`.
var	numeric vector containing the two variables to be plotted on the axis. Used only if `type='right'`.
zoomed	if `TRUE`, shows only the area where most of the density is concentrated. If `FALSE`, shows the whole area.
showgrid	if `TRUE`, shows the axis and the labels. If `FALSE`, only the graph is printed.
nlevels	approximate number of contour lines to be drawn.

Details

The number of variables in the Flexible Dirichlet must be 3 to draw a plot. Vectors a and p must be of the same length.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

[FD.generate](), [FD.density]()

Examples

```
alpha <- c(12,7,15)
prob <- c(0.3,0.4,0.3)
tau <- 8
FD.theorcontours(alpha,prob,tau)
FD.theorcontours(alpha,prob,tau, type='right', var=c(3,2), zoomed=FALSE, showgrid=TRUE, nlevels=3)
```

---

oliveoil *Olive oil data*

Description

This data set represents eight chemical measurements on different specimen of olive oil produced in various regions in Italy (northern Apulia, southern Apulia, Calabria, Sicily, inland Sardinia and coast Sardinia, eastern and western Liguria, Umbria) and further classifiable into three macro-areas: Centre-North, South, Sardinia.

Format

This data frame contains 572 rows, each corresponding to a different specimen of olive oil, and 10 columns. The first and the second column correspond to the macro-area and the region of origin of the olive oils respectively; here, the term 'region' refers to a geographical area and only partially to administrative borders. Columns 3-10 represent the following eight chemical measurements on the acid components for the oil specimens: palmitic, palmitoleic, stearic, oleic, linoleic, linolenic, arachidic, eicosenoic.

Source

Originally included in the package pdfCluster.

---

plot.FDfitted *Plot Method for FDfitted Objects*

Description

This method plots the results of FD.estimation, using the functions FD.ternaryplot or FD.rightplot.

Usage

```
## S3 method for class 'FDfitted'
plot(x, type = "ternary", var = c(1, 2), zoomed = T,
  showgrid = T, showdata = T, nlevels = 10, ...)
```

Arguments

x	an object of class FDfitted, usually the result of `FD.estimation`.
type	string containing `'ternary'` or `'right'`.
var	numeric vector containing the two variables to be plotted on the axis. Used only if `type='right'`.
zoomed	if TRUE, shows only the area where most of the density is concentrated. If FALSE, shows the whole area.
showgrid	if TRUE, shows the axis and the labels. If FALSE, only the graph is printed.
showdata	if TRUE, prints the data points. If FALSE, shows only the contour lines.
nlevels	approximate number of contour lines to be drawn.
...	additional arguments

Details

The number of variables in the fitted model must be 3 to draw a plot.

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

`FD.estimation`, `FD.ternaryplot`, `FD.rightplot`, `FD.marginalplot`

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
plot(results)
plot(results, type='right', var=c(3,2), zoomed=FALSE, showgrid=TRUE, showdata=FALSE, nlevels=3)
```

---

print.FDfitted   *Print Method for FDfitted Objects*

Description

This method shows the results of `FD.estimation`.

Usage

```
## S3 method for class 'FDfitted'
print(x, ...)
```

Arguments

x	an object of class FDfitted, usually the result of `FD.estimation`.
...	additional arguments

---

summary.FDfitted   *Summary Method for FDfitted Objects*

Description

This method summarizes the results of `FD.estimation`, adding also information from the functions `FD.stddev` and `FD.aicbic`.

Usage

```
## S3 method for class 'FDfitted'
summary(object, ...)
```

Arguments

object	an object of class FDfitted, usually the result of `FD.estimation`.
...	additional arguments

Value

A list composed by:

par Estimated parameter vector

sd Vector of the estimated standard deviations

goodness Vector containing LogLikelihood, AIC and BIC

References

Ongaro, A. and Migliorati, S. (2013) A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, **114**, 412–426.

Migliorati, S., Ongaro, A. and Monti, G. S. (2016) A structured Dirichlet mixture model for compositional data: inferential and applicative issues. Statistics and Computing, doi:10.1007/s11222-016-9665-y.

See Also

`FD.estimation`, `FD.stddev`, `FD.aicbic`

Examples

```
data <- FD.generate(n=20,a=c(12,7,15),p=c(0.3,0.4,0.3),t=8)
data
results <- FD.estimation(data, normalize=TRUE,iter.initial.SEM = 5,iter.final.EM = 10)
results
summary(results)
```

Index