Package ‘MDM’

February 19, 2015

Type Package
Title Multinomial Diversity Model
Version 1.3
Date 2013-06-28
Author Glenn De'ath <g.death@aims.gov.au>; Code for mdm was adapted from multinom in the nnet package (Brian Ripley <ripley@stats.ox.ac.uk> and Bill Venables <Bill.Venables@csiro.au>)
Maintainer Glenn De'ath <g.death@aims.gov.au>
Depends R (>= 2.9.0), nnet
Description The multinomial diversity model is a toolbox for relating diversity to complex predictors. It is based on (1) Shannon diversity; (2) the multinomial logit model, and (3) the link between Shannon diversity and the log-likelihood of the MLM.
License GPL-2 | GPL-3
Repository CRAN
Date/Publication 2013-07-10 11:12:01
NeedsCompilation no

R topics documented:

MDM-package .................................................. 2
anova.mdm .................................................. 2
dev2div ...................................................... 4
ed ......................................................... 5
del ......................................................... 6
eds .......................................................... 6
eds1 ........................................................... 7
entropy.plot ................................................... 8
mdm .......................................................... 9
simdata ...................................................... 11
spider6 ..................................................... 13
y2p .......................................................... 14

Index 15
Description

The multinomial diversity model is a toolbox for relating diversity to complex predictors. It is based on (1) Shannon diversity; (2) the multinomial logit model, and (3) the link between Shannon diversity and the log-likelihood of the MLM.

Details

Package: MDM
Type: Package
Version: 1.0
Date: 2011-09-08
License: GPL (version 2 or newer)
LazyLoad: yes

Author(s)

Glenn De’ath: &lt;g.death@aims.gov.au&gt;

References


Examples

library(MDM)
data(spider6)
fit0 &lt; mdm(y2p(spider6[,1:6])~1,data=spider6)
fit1 &lt; mdm(y2p(spider6[,1:6])~Water,data=spider6)
fit2 &lt; mdm(y2p(spider6[,1:6])~Water+Herbs,data=spider6)
fit3 &lt; mdm(y2p(spider6[,1:6])~Site,data=spider6,alpha=TRUE)
anova(fit0,fit1,fit2,fit3)

Analysis of Deviance, Entropy and Diversity Tables
Description

Provides an analysis of deviance, entropy and diversity for a collection of diversity models (outputs from `mdm`).

Usage

```r
## S3 method for class 'mdm'
anova(object, ..., topnote = TRUE, cols = c("df","dev","ent","div")[1:4])
```

Arguments

- `object, ...`: objects of class `mdm`, usually, a result of a call to `mdm`.
- `topnote`: If TRUE then model descriptions appear above the anova table, if FALSE they appear as the first column of the table.
- `cols`: The list of columns to print out. Defaults to all columns.

Details

Specifying a single object gives a sequential analysis of deviance table for that fit. That is, the reductions in the residual sum of deviances, plus the residual deviance. The deviances are converted into entropies and diversities, and differences in deviances are converted into differences in entropies and diversities.

If more than one object is specified, the table has a row for the residual degrees of freedom and sum of deviances for each model. For all but the first model, the change in degrees of freedom and deviances. This only makes statistical sense if the models are nested. It is conventional to list the models from smallest to largest, but this is up to the user.

Value

The analysis of deviance, entropy and diversity for a collection of diversity models.

Note

`mdm` is a modified version of `multinom` in the `nnet` package.

References


See Also

`multinom, nnet`
Examples

```r
library(MDM)
data(spider6)
fit0 <- mdm(y2p(spider6[,1:6])-1, data=spider6)
fit1 <- mdm(y2p(spider6[,1:6])-Water, data=spider6)
fit2 <- mdm(y2p(spider6[,1:6])-Water+Herbs, data=spider6)
fit3 <- mdm(y2p(spider6[,1:6])-Site, data=spider6)
anova(fit0, fit1, fit2, fit3)
```

---

**dev2div**

*Converts deviances to diversities*

---

**Description**

Takes either (1) a `mdm` or (2) a scaler, vector or matrix of deviances, and extracts or converts them to diversities. The relationship between deviance (dev) and diversity (d) is given by \( d = \exp({dev}/2/n) \) where \( n \) is the number of units (typically rows of a matrix) over which deviance is being averaged.

**Usage**

```r
dev2div(x, n)
```

**Arguments**

- `x`  
  A `mdm` or a scaler, vector or matrix of deviances.

- `n`  
  If `x` is not a `mdm`, then the divisor in the conversion as defined as above.

**Value**

The diversity of \( x \).

**See Also**

`ed`, `ed1`, `eds`, `eds1`

**Examples**

```r
x <- c(5,10,15)
dev2div(x, n=10)
```
Calculate alpha, beta and gamma true entropies and diversities

**Description**

Calculates alpha, beta and gamma true diversities of a data set. The matrix or data frame is automatically scaled to row sums of one.

**Usage**

```r
ed(x, q = 1, w = 1, retq = TRUE)
```

**Arguments**

- `x` the input matrix or data frame.
- `q` the order of diversity; typically 0, 1 or 2.
- `w` weights if required.
- `retq` if TRUE then diversities are returned; if FALSE the entropies for alpha and gamma are returned.

**Value**

a vector of entropies or diversities

**See Also**

`dev2div`, `ed1`, `eds`, `eds1`

**Examples**

```r
data(spider6)
ed(spider6[,1:6])
ed(spider6[,1:6], q=0)
ed(spider6[,1:6], q=2)
ed(spider6[,1:6], retq=FALSE)
data(spider6)
ed(spider6[,1:6])
```
Calculate alpha, beta and gamma true entropies and diversities

Description
Calculates true diversities of individual sites of a data set

Usage
```
ed1(x, q = 1, retq = TRUE)
```

Arguments
- `x`: the input matrix or data frame.
- `q`: the order of diversity; typically 0, 1 or 2.
- `retq`: if TRUE then diversities are returned; if FALSE the entropies for alpha and gamma are returned.

Value
a vector of entropies or diversities

See Also
`dev2div`, `ed1`, `eds`, `eds1`

Examples
```
data(spider6)
ed1(spider6[,1:6])
ed1(spider6[,1:6],q=0)
ed1(spider6[,1:6],q=2)
ed1(spider6[,1:6],retq=FALSE)
data(spider6)
ed1(spider6[,1:6])
```

Calculate alpha, beta and gamma parametric entropies and diversities

Description
Calculates alpha, beta and gamma parametric diversities of a data set. The matrix or data frame is automatically scaled to row sums of one. Unlike true diversities where weighting is done through the function ed, weighting for parameterized diversity is done within the MDM, or more generally by using case weights.
**Usage**

```
eds(x, q = 1, w = 1, retq = TRUE)
```

**Arguments**

- `x` the input matrix or data frame.
- `w` weights if required.
- `q` the order of diversity; typically 0, 1 or 2.
- `retq` if TRUE then parametric diversities are returned; if FALSE the entropies for alpha and gamma are returned.

**Value**

a vector of entropies or diversities

**See Also**

`dev2div, ed1, ed, eds1`

**Examples**

```r
data(spider6)
eds(spider6[,1:6])
eds(spider6[,1:6],q=0)
eds(spider6[,1:6],q=2)
eds(spider6[,1:6],retq=FALSE)
data(spider6)
eds(spider6[,1:6])
```

---

**eds1**

*Calculate alpha, beta and gamma parametric entropies and diversities*

**Description**

Calculates parametric diversities of individual sites of a data set

**Usage**

```
eds1(x, q = 1, retq = TRUE)
```

**Arguments**

- `x` the input matrix or data frame.
- `q` the order of diversity; typically 0, 1 or 2.
- `retq` if TRUE then parametric diversities are returned; if FALSE the entropies for alpha and gamma are returned.
Value

a vector of entropies or diversities

See Also

dev2div, ed1, eds, eds1

Examples

data(spider6)
edls(spider6[,1:6])
edls(spider6[,1:6],q=0)
edls(spider6[,1:6],q=2)
edls(spider6[,1:6],retq=FALSE)
data(spider6)
edls(spider6[,1:6])

entropy.plot

Description

Plots components of entropy by species or sites.

Usage

entropy.plot(lst, y, x, ord=TRUE, type=c("species","sites"), labs, segs = TRUE, widseg=0.25, pchs = c(15,19,0,2,6,2,15,17,21,17), ...)

Arguments

lst a list of mdm fitted models
y optional – locations of the components on the vertical axis
x optional – used to provide the range of the horizontal axis
ord if TRUE orders the components by total entropy of each fitted model.
type should the plot break down entropy by species (columns = default) or sites (rows).
labs an (optional) vector of labels for species (or sites if type == "sites").
segs use segments to plot end of bars as vertical lines
widseg widths of bar ends
pchs some nice plotting characters
... other arguments passed to the plot function (e.g. cex).
mdm

Value

a list of plotted values of the length of lst, each component of which is of length equal to the number of species (sites).

See Also

mdm

Examples

data(spider6)
fit0 <- mdm(y2p(spider6[,1:6])~1, data=spider6)
fit1 <- mdm(y2p(spider6[,1:6])~Water+Herbs, data=spider6)
fit2 <- mdm(y2p(spider6[,1:6])~Site, data=spider6, alpha=TRUE)
anova(fit0, fit1, fit2)
entropy.plot(list(fit0, fit2, fit1))

Description

The parametric diversity model (mdm) is a new method for directly relating diversity to environmental predictors. It is based on three components: (1) parametric diversity, a new parametric model of diversity that can represent any configuration of species proportional abundances, (2) the multinomial logit model (MLM) that can relate species proportional abundances to complex predictors and (3) the link between parametric diversity and the likelihood function of the MLM.

The mdm is fitted using the multinom package from nnet. Parametric diversities and true diversities can also be calculated for data sets using the functions eds, eds1, ed, ed1.

Usage

mdm(formula, data, weights, subset, na.action,
    MaxNWts, maxit = 1000, contrasts = NULL, Hess = FALSE,
    censored = FALSE, model = TRUE, use.shortcut = TRUE, ...)

Arguments

formula a formula expression as for regression models, of the form response ~ predictors. The response should be a matrix with K columns comprising proportions for each of K classes. A log-linear model is fitted, with coefficients zero for the first class. An offset can be included: it should be a numeric matrix with K columns. See the documentation of formula() for other details.

data an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which mdm is called.
weights | optional case weights in fitting.
subset | expression saying which subset of the rows of the data should be used in the fit. All observations are included by default.
na.action | a function to filter missing data.
MaxNWts | The maximum allowable number of weights. There is no limit in the code, but MaxNWts is set to the exact number of required as specified in the formula. Thus it should not need to be changed when fitting `mdm`.
maxit | maximum number of iterations. Default 1000.
contrasts | a list of contrasts to be used for some or all of the factors appearing as variables in the model formula.
Hess | logical for whether the Hessian (the O/E information matrix) should be returned.
censored | If Y is a matrix with K > 2 columns, interpret the entries as one for possible classes, zero for impossible classes, rather than as counts.
model | logical. If true, the model frame is saved as component model of the returned object.
useshortcut | logical. If true, and the model is ~1 (a constant) or ~sites (a factor with one level for each site) then the model is not fitted since the fitted values are known in each case. The first (alpha) model has fitted values equal to the input data and the second (gamma) model fits the row means. Fitting the alpha-model using nnet can be prohibitively expensive in computational time and is unnecessary. The returned models when useshortcut == TRUE has the same components as when useshortcut == FALSE, and hence can be used in anova tables and plotting. Using useshortcut == TRUE can result in saving > 99% of computational time for a collection of models.
... | additional arguments for nnet.

Details

`mdm` calls `multinom` which calls `nnet`. The variables on the rhs of the formula should be roughly scaled to [0,1] or the fit will be slow or may not converge at all.

Value

A nnet object with additional components:

deviance | the residual deviance, compared to the full saturated model (that explains individual observations exactly). Also, minus twice log-likelihood
edf | the (effective) number of degrees of freedom used by the model
AIC | the AIC for this fit
Hessian | if Hess is true
model | if model is true
entropy | the entropy of the fitted values
diversity | the diversity of the fitted values
**simdata**

*Species abundance data simulator*

**Description**

Simulates species abundance data along a one-dimensional gradient

**Usage**

```r
simdata(d, p = 10, n = 100, strip0 = TRUE, extreme = 0,
ret = TRUE, k.rand = FALSE, d.rand = TRUE, mu.rand = TRUE,
s = rep((4:8)/10, length = p), amp = c(sample(1:5), rep = TRUE,
length = p), skew = 1, ampfun, lst = FALSE, err = 1,
err.type = c("p", "n")[1], as.df = TRUE, plotit = TRUE,
ptype = "l", plty = 1, pcols = rainbow(p), add.rug = FALSE, ...)```

**Note**

`mdm` is a modified version of `multinom` in the `nnet` package.

**References**


**See Also**

`multinom`, `nnet`
Arguments

- **d**: the (optional) locations of the species along the 1-D gradient. If d is given then it will define both the number of species and also the locations on the gradient e.g. \(d = \text{rep}(1:10,\text{each}=3)\) will generate species at locations 1,1,2,2,2,...,10,10,10. If d is not specified then d.rand = TRUE will randomly allocate the species modes along a gradient on \([0, 1]\), but if d.rand = FALSE will uniformly distribute the species modes along a gradient.

- **p**: number of species.

- **n**: number of sites.

- **stripP**: if TRUE the sites with zero total abundance are omitted.

- **extreme**: number typically in the range -1 to +1 with larger numbers reducing the range of species.

- **ret**: if TRUE the generated data are returned

- **k.rand**: should the be random (TRUE) or fixed

- **d.rand**: should the be random (TRUE) or fixed

- **mu.rand**: should the be random (TRUE) or fixed

- **s**: the spans of the species response curves; s is the standard deviation of the spread

- **amp**: the amplitudes of the species response curves

- **skew**: the skewness of the distribution; range (>0 to 5), 1 = symmetric.

- **ampfun**: any function to modify the amplitude

- **lst**: if lst == TRUE then both the systematic and random values are returned

- **err**: if err == 0 then the values are systematic with no random variation

- **err.type**: type of error; p = poison, g = gaussian

- **as.df**: if return returns a data frame, otherwise a matrix

- **plotit**: if TRUE then the data are plotted

- **ptype**: species plot types e.g. "l" gives lines

- **plty**: species plot line types

- **pcols**: species plot colours

- **add.rug**: should a rug be added?

- **...**: other arguments passed to plot.

Value

If lst == FALSE then a data frame with variables "Locations", "Taxa.1" – "Taxa.N" where N is number of species. If lst == TRUE then two data frames "x" and "xs" with variables "Locations", "Taxa.1" – "Taxa.N" and additionally, components "sigma", "amp" and "mu" that represent the spans, amplitudes and locations of the N species along the 1-D gradient.

Examples

```r
mydata <- simdata()
summary(mydata)
mydata <- simdata(p=5, n=50, amp=1, err=0, d.rand=FALSE,
                  mu.rand=FALSE, plotit = TRUE)
summary(mydata)
```
### The spider data set

#### Description
Data set on abundances of spiders and environmental predictors. This is a subset of a larger data comprising 12 species and 6 environmental predictors. All variables are rated on a 0-9 scale.

#### Usage
```r
data(spider6)
```

#### Format
A data frame with 28 observations on the following 9 variables.

- `Pard.lugu` a numeric vector
- `Pard.pull` a numeric vector
- `Troc.terr` a numeric vector
- `Pard.mont` a numeric vector
- `Alop.acce` a numeric vector
- `Alop.fabr` a numeric vector
- `Water` a numeric vector
- `Herbs` a numeric vector
- `Site` a factor with 28 levels

#### Source
package mvpart

#### References

#### Examples
```r
data(spider6)
summary(spider6)
fit0 <- mdm(y2p(spider6[,1:6])~1,data=spider6)
fit1 <- mdm(y2p(spider6[,1:6])~Water+Herbs,data=spider6)
fit2 <- mdm(y2p(spider6[,1:6])~Site,data=spider6,alpha=TRUE)
anova(fit0,fit1,fit2)
```
Description

The response matrix for mdm diversity analyses requires that species (columns of the data matrix) are scaled to proportions i.e. that they sum to one for each site (row of the data matrix). Two matrices are particularly useful; these correspond to alpha and gamma diversities. The former is particularly useful for mdm anovas that include the alpha model which is computationally expensive using the function mdm.

Usage

\[ y2p(y, \text{mean}=\text{FALSE}) \]

Arguments

- **y**: matrix or data frame of numeric values to be transformed
- **mean**: if mean = TRUE then each row is replaced by the species means scaled to row sums of one

Value

A matrix of the same dimensions as the input matrix. Each row of the matrix will sum to one.

Examples

```r
mydata <- matrix(0:8,nrow=3,ncol=3)
mydata
y2p(mydata)
```
Index

+Topic **datasets**
  spider6, 13
+Topic **diversity**
  MDM-package, 2
+Topic **mdm**
  anova.mdm, 2
  mdm, 9
  anova.mdm, 2
  dev2div, 4, 5–8
  ed, 4, 5, 7, 9
  ed1, 4–6, 6, 7–9
  eds, 4–6, 6, 8, 9
  eds1, 4–7, 7, 8, 9
  entropy.plot, 8

  MDM (MDM-package), 2
  mdm, 3, 8, 9, 9, 10, 11, 14
  MDM-package, 2
  multinom, 3, 10, 11

  nnet, 3, 10, 11

  simdata, 11
  spider6, 13

  y2p, 14