Package ‘mudfold’

December 18, 2017

Type Package

Title Multiple UniDimensional unFOLDing

Version 1.1.0

Date 2017-12-11

Description Nonparametric item response theory (IRT) model for nonmonotonic IRFs, which is fruitful for scale analysis of proximity items.

Depends R (>= 3.3.3), yesno, gtools, zoo, ggplot2, reshape2, stats, utils

License GPL (>= 2)

URL http://www.math.rug.nl/stat/Main/HomePage

Maintainer Spyros Balafas <s.balafas@rug.nl>

Repository CRAN

NeedsCompilation no

Author Spyros Balafas [aut, cre],
Wim Krijnen [aut],
Wendy Post [ctr],
Ernst Wit [aut]

Date/Publication 2017-12-18 14:04:10 UTC

R topics documented:

mudfold-package 2
ANDRICH 3
as.mudfold 4
EURPAR2 7
mudfold 8
mudfoldsim 13
pick 15
Plato7 17
plot.mdf 18
print.mdf 19
summary.mdf 20

Index **22**

`mudfold-package` *MUDFOLD : A nonmonotonic nonparametric model for undolding scale analysis.*

Description

This package can be used for the purpose of finding unfolding structures from selected items in tests or questionnaires. Such structures, represent the underlying latent ordering of those items. The main function of this package is called `mudfold` and fits the Van Schuur's scale construction method to binary proximity items. The method is called Multiple UniDimensional unFOLDing (MUDFOLD) and is an item selection algorithm belonging in the class of Nonparametric Item Response Theory (IRT) models.

Details

MUDFOLD is a nonparametric probabilistic model for unidimensional unfolding. Originally developed by W. Van Schuur (1984) and further extended following ideas by W.J. Post (1992) who derived testable properties for the model fit. This method can be used to analyse proximity items pressumably generated from a nonmonotonic (unimodal) Item Response Function (IRF). The package incorporates the main function `mudfold` which is used to estimate the MUDFOLD scale to binary valued proximity items. The output of the main function is a list of S3 class `.mdf`, for which `print()`, `summary()` and `plot()` generic functions are available to the user.

The data must be given in an $n \times p$ binary `matrix` or `data.frame` with n respondents in the rows and p items in the columns. Each row of the data corresponds to the selections of the i-th individual on a set of p items. Missing values (NA) are not allowed.

Ultimate goal for MUDFOLD is to determine a unidimensional rank order of a (sub)set of items such that, they constitute an appropriate scale for measuring a common latent trait of the respondents. The estimation of the item order is done through an heuristic item selection algorithm, which tests iteratively the item fit to the scale with the use of scalability coefficients.

MUDFOLD's H coefficients of scalability are based to Loevinger's coefficient of homogeneity. In MUDFOLD, H coefficients utilize a scalability measure that is used in several criteria in the item selection algorithm. This coefficient in MUDFOLD can be calculated for triples of items, individual items, and the total scale.

Author(s)

Spyros E. Balafas (auth.), Wim P. Krijnen (auth.), Wendy J. Post (contr.), Ernst C. Wit (auth.)

Maintainer: Spyros E. Balafas (<s.balafas@rug.nl>)

References

W.H. Van Schuur.(1984). *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. CT Press.

W.J. Post. (1992). *Nonparametric Unfolding Models: A Latent Structure Approach*. M & T series. DSWO Press.

W.J. Post. and T.AB. Snijders (1993). *Nonparametric unfolding models for dichotomous data*. Methodika.

Examples

```
## Not run:
# Install the R package mudfold
install.packages("mudfold")

# Load the R package mudfold
library(mudfold)


## End(Not run)
```

ANDRICH *Andrich's attitude scale towards capital punishment*

Description

D. Andrich's (1988) scale designed to measure the attitude from a sample of students towards capital punishment. The data set contains the dichotomous responses of 54 students on 8 statements concerning capital punishment.

Usage

```
data(ANDRICH)
```

Format

A data frame with 54 observations on the following 8 variables.

HIDEOUS a numeric vector containing the binary responses on the statement:
"Capital punishment is one of the most hideous practices of our time"

LIFESACRED a numeric vector containing the binary responses on the statement:
"The state cannot teach the sacredness of human life by destroying it"

INEFFECTIV a numeric vector containing the binary responses on the statement:
"Capital punishment is not an effective deterrent to crime"

DONTBELIEV a numeric vector containing the binary responses on the statement:
"I do not believe in capital punishment but i am not sure it is not necessary"

WISHNOTNEC a numeric vector containing the binary responses on the statement:
"I think capital punishment is necessary but i wish it were not"

MUSTHAVEIT a numeric vector containing the binary responses on the statement:
"Until we find a more civilized way to prevent crime we must have capital punishment"

DETERRENT a numeric vector containing the binary responses on the statement:
"Capital punishment is justified because it does act as a deterrent to crime"

CRIMDESERV a numeric vector containing the binary responses on the statement:
"Capital punishment gives the criminal what he deserves"

Details

The persons who responded to the statements for the analysis were 54 graduate students taking an introductory course in educational measurement and statistics. They responded simply by agreeing (1) or disagreeing (0) with each statement, with no restrictions placed on how many statements should receive an *Agree* response.

Source

D. Andrich. (1988). *The Application of an Unfolding Model of the PIRT Type to the Measurement of Attitude*. Applied psychological measurement 12.1 (1988): 33-51.

References

D. Andrich. (1988). *The Application of an Unfolding Model of the PIRT Type to the Measurement of Attitude*. Applied psychological measurement 12.1 (1988): 33-51.

W.J. Post. (1992). *Nonparametric Unfolding Models: A Latent Structure Approach*. M & T series. DSWO Press.

W.J. Post. and T.AB. Snijders. (1993). *Nonparametric unfolding models for dichotomous data*. Methodika.

Examples

```
## Not run:
data(ANDRICH)
str(ANDRICH)

## End(Not run)
```

---

as.mudfold — *Function for calculating MUDFOLD statistics for a given unfolding scale.*

Description

This function calculates the MUDFOLD statistics for proximity item response data, whose columns are assumed to be ordered to the order they are provided. The resulting object from the `as.mudfold` function is an object of S3 class `.mdf`, for which generic functions `print`, `summary`, and `plot` are available.

Usage

```
as.mudfold(x,estimation="rank")
```

Arguments

- `x` : A binary matrix or data frame containing the responses of `nrow(x)` persons to `ncol(x)` items. Missing values in `x` are not allowed.
- `estimation` : This argument controls the nonparametric estimation method for item and subject locations. By deafult this argument equals to `"rank"` and implies that a Van Schuur's rank based estimator will be used for estimating the item parameters which are later used as thresholds in order to estimate subject 's parameters. The user can set this argument to `"quantile"` and then an estimator based on item rank quantiles proposed by Johnson is applied.

Details

The function `as.mudfold` calculates MUDFOLD statistics for a given scale. Descriptive statistics, scalability coefficients, iso statistic values, are calculated and the user can obtain a summary table with the `summary` function which is designed for `.mdf` class objects.

Value

The function `as.mudfold` returns a list object of class `mdf` with the same components as the core function `mudfold` except the information that concerns the item selection algorithm. The list contains the following:

- `dat` The data in which MUDFOLD method has been fitted.
- `starting.items` The starting set of items.
- `no.items` The number of items in the starting set (i.e., equal to `ncol(x)`).
- `sample.size` The number of respondents.
- `Best.triple` `NA`
- `iterations.in.sec.step` `NA`
- `mdfld.order` The item order on the unidimensional scale.
- `length.scale` The number included in the scale (`length(mdfld.order)`).
- `item.popularities` Observed proportion of positive responses for the items included in the scale.
- `item.freq` Observed frequency of positive responses for the items in MUDFOLD scale.
- `Obs.err.item` Observed response errors for each item included in the scale.
- `Exp.err.item` Expected response errors for each item included in the scale.
- `H.item` Scalability coefficient for each item included in the scale.
- `Item.ISO` Iso statistic for each item included in the scale.
- `Obs.err.scale` Observed response errors for the estimated scale.
- `Exp.err.scale` Expected response errors for the estimated scale
- `Htotal` Scalability coefficient for the estimated scale.
- `Isototal` Iso statistic for the estimated scale.
- `Cond.Adjacency.matrix` Conditional adjacency matrix (CAM) for the estimated scale.

`Adjacency.matrix`
: Adjacency matrix for the estimated scale.

`Dominance.matrix`
: Dominance matrix for the MUDFOLD scale.

`star`
: A matrix with stars placed at the maxima locations of each row of the conditional adjacency matrix.

`Correlation.matrix`
: Correlation matrix for the MUDFOLD scale.

`uniq`
: The set of unique triples. From this set, the best minimal scale for the first step of the item selection algorithm is determined.

`est.parameters`
: A list with two components. The first component refers to item parameters and the other to the subject parameters. The estimates have been obtained with a user specified nonparametric estimation method.

`call`
: The function call.

Author(s)

Spyros E. Balafas (auth.), Wim P. Krijnen (auth.), Wendy J. Post (contr.), Ernst C. Wit (auth.)

Maintainer: Spyros E. Balafas (<s.balafas@rug.nl>)

References

W.H. Van Schuur.(1984). *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. CT Press.

W.J. Post. (1992). *Nonparametric Unfolding Models: A Latent Structure Approach*. M & T series. DSWO Press.

W.J. Post. and T.AB. Snijders. (1993).Nonparametric unfolding models for dichotomous data. Methodika.

M.S. Johnson. (2006). Nonparametric Estimation of Item and Respondent Locations from Unfolding-type Items. Psychometrica

See Also

`mudfold`

Examples

```
## Not run:
## pick a number for setting the seed
n.seed <- 11

## Simulate an unfolding scale
simulation <- mudfoldsim(p=6, n=100, gamma1=5, gamma2=-10, zeros=FALSE,seed=n.seed)

## get the data
dat <- simulation$dat
```

```
## true order
true_order <- simulation$true_ord

## check MUDFOLD statistics for the simulated order
mud_stats1 <- as.mudfold(dat,estimation="rank")

# get the summary table
summary(mud_stats1)

## check MUDFOLD statistics for the true item order
mud_stats2 <- as.mudfold(dat[,true_order],estimation="rank")

# get the summary table
summary(mud_stats2)

## End(Not run)
```

`EURPAR2` *Preferences of European party activists.*

Description

European party activists preferences for two political parties in the European parliament in 1980. A sample consisted of 1786 individuals are asked to pick 2 out of 6 political parties from the European parliament.

Usage

```
data("EURPAR2")
```

Format

A data frame with 1786 observations (responses) on the following 6 variables (items).

`communists` Communistic political party;

`socdemocr` Social Democratic political party;

`demprogres` Progressive Democratic political party;

`liberals` Liberal Democratic political party;

`christians` Christian Democratic political party;

`conservat` Conservative political party;

Details

The data have been first studied by Van Schuur (1984) and further by W. J. Post (1992).

Source

W.H. Van Schuur.(1984). *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. CT Press.

References

W.H. Van Schuur.(1984). *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. CT Press.

W.J. Post. (1992). *Nonparametric Unfolding Models: A Latent Structure Approach*. M & T series. DSWO Press.

Examples

```
data(EURPAR2)
str(EURPAR2)
```

---

mudfold — *MUDFOLD item selection algorithm for dichotomous proximity items.*

Description

This function incorporates a two step algorithm that determines an unfolding scale from observed binary data `x`. In the first step of the algorithm the best scale consisting of `3` items is determined, while in the second step, the scale obtained in the first step is expanded iteratively by adding the best fitting item in each iteration. The first step of the algorithm can be skiped with the argument `start` which can be used by setting manually an ordered item set that will be extended in the second step of the item selection algorithm. The resulting scale consists of the best `m` fitting items based on scalability criteria (where `m` $\leq$ `ncol(x)`).

In `mudfold` function, the user can specify a value λ_1 that will be used as a lower boundary in the scalability criteria of the search algorithm. By default, the lower boundary for the scalability coefficients is `lambda1=0.3` (Mokken,1971). The user can choose a second value λ_2 (usually negative) that will be used as a lower boundary only for the second step of the algorithm (by default, `lambda2=0`). The parameter λ_2 is used mostly, in order to relax the first scalability criterion of the second step. Generally, values greater than 0.3 for λ_1, and λ_2 lead to very strict criteria while negative values relax these criteria.

Moreover, the user is able to choose between two nonparametric estimation methods in order to obtain individual and item scale parameters that are estimated based on the resulting item order. The default setting (i.e., `estimation="rank"`) uses an estimation proposed by Van Schuur based on item ranks. Alternatively, an estimation method described by M.S. Johnson, which uses item quantiles for estimating person parameters, can be used by setting `estimation="quantile"`.

Usage

```
mudfold(x, estimation="rank", lambda1=.3, lambda2=0,start= NULL,check=TRUE)
```

Arguments

`x`	: A binary matrix or data frame containing the responses of `nrow(x)` persons to `ncol(x)` items. Missing values in `x` are not allowed.

`estimation` : This argument controls the nonparametric estimation method for item and subject locations. By deafult this argument equals to `"rank"` and implies that Van Schuur's rank based estimator will be used for estimating the item parameters which are later used as thresholds in order to estimate subject 's parameters. The user can set this argument to `"quantile"` and then an estimator based on item rank quantiles proposed by Johnson is applied.

`lambda1` : User specified numerical value that is used as a lower boundary for the scalability criterion of the first step of the item selection algorithm, and in the item scalability criterion at the end of the scale expansion. Default value is $\lambda_1 = 0.3$ but it can be any value between $-\infty$ and 1 (i.e., $\lambda_1 \in (-\infty, 1]$). The higher the value of λ_1 the stricter the scalability criteria of the algorithm.

`lambda2` : User specified numerical value that controls explicitly the first scalability criterion of the scale expansion. In the default settings $\lambda_2 = 0$, however, the user can choose a negative value for λ_2, which leads to less strict scalability criterion in the beginning of the scale expansion.

`start` : A pressumably ordered character vector containing column names of `x`, with length greater than or equal to 3 and less than or equal to the number of columns of `x`. This ordered item set is used as a startset for the scale extension phase of MUDFOLD method. If `start= NULL` the standard MUDFOLD method is fitted to the data.

`check` : A logical argument. If `check=TRUE` (default) then the data is checked for errors.

Details

In the first stage, `mudfold` function seeks the best ordered triple of items. The best triple of items is called *best unique triple* and has to fullfil certain scalability criteria in order to be chosen as the best elementary scale. Specifically, the scalability coefficient of the best unique triple has to be higher than a user specified parameter λ_1.

In the second stage of the algorithm the best elementary scale is extended by adding the best fitting item in each replication of the algorithm based again on scalability criteria. The user can potentially specify an explicit lower boundary for the first scalability criterion of the second step through a parameter λ_2.

The parameters λ_1, λ_2 can take any value in $(-\infty, 1]$. Values larger than 0.3 for the λ's lead to very strict scalability criteria.

Value

The function `mudfold` returns a list object of class `mdf` with the following components:

`dat` The data in which MUDFOLD method has been fitted.

`starting.items` The starting set of items.

`no.items` The number of items in the starting set (i.e., equal to `ncol(data)`).

`sample.size` The number of respondents.

`Best.triple` The best minimal scale (triple of items) determined in the first step of the item selection algorithm.

iterations.in.sec.step
: Number of iterations in the second step of MUDFOLD method.

mdfld.order
: The item order resulted from the item selection algorithm.

length.scale
: The number included in the scale (`length(mdfld.order)`).

item.popularities
: Observed proportion of positive responses for the items included in the scale.

item.freq
: Observed frequency of positive responses for the items in MUDFOLD scale.

Obs.err.item
: Observed response errors for each item included in the scale.

Exp.err.item
: Expected response errors for each item included in the scale.

H.item
: Scalability coefficient for each item included in the scale.

Item.ISO
: Iso statistic for each item included in the scale..

Obs.err.scale
: Observed response errors for the estimated scale.

Exp.err.scale
: Expected response errors for the estimated scale

Htotal
: Scalability coefficient for the estimated scale.

Isototal
: Iso statistic for the estimated scale.

Cond.Adjacency.matrix
: Conditional adjacency matrix (CAM) for the estimated scale.

Adjacency.matrix
: Adjacency matrix for the estimated scale.

Dominance.matrix
: Dominance matrix for the MUDFOLD scale.

star
: A matrix with stars placed at the maxima locations of each row of the conditional adjacency matrix.

Correlation.matrix
: Correlation matrix for the MUDFOLD scale.

uniq
: The set of unique triples. From this set, the best minimal scale for the first step of the item selection algorithm is determined.

est.parameters
: A list with two components. The first component refers to item parameters and the other to the subject parameters. The estimates have been obtained with a user specified nonparametric estimation method.

call
: The function call.

Author(s)

Spyros E. Balafas (auth.), Wim P. Krijnen (auth.), Wendy J. Post (contr.), Ernst C. Wit (auth.)

Maintainer: Spyros E. Balafas (<s.balafas@rug.nl>)

References

Mokken, R. J. (1971). A theory and procedure of scale analysis: With applications in political research (Vol. 1). Walter de Gruyter.

W.H. Van Schuur.(1984). *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. CT Press.

W.J. Post. (1992). *Nonparametric Unfolding Models: A Latent Structure Approach*. M & T series. DSWO Press.

W.J. Post. and T.AB. Snijders. (1993).Nonparametric unfolding models for dichotomous data. Methodika.

M.S. Johnson. (2006). Nonparametric Estimation of Item and Respondent Locations from Unfolding-type Items. Psychometrica

Examples

```
## Not run:
#####################################
#### MUDFOLD method on real data ####
#####################################

### MUDFOLD method on ANDRICH data (see Post and Snijders pp.147) ###
data(ANDRICH)

## fit MUDFOLD on ANDRICH data ##
fit_andr <- mudfold(ANDRICH)

## generic functions for the S3 class .mdf object fit ##

# print
fit_andr

# summary
summary(fit_andr)

# plot
plot(fit_andr)

### MUDFOLD method on EURPAR2 data ###
data("EURPAR2")

## fit MUDFOLD on EURPAR2 data ##
fit_eurp <- mudfold(EURPAR2)

# print
print(fit_eurp,Diagnostics = TRUE)

# summary
summary(fit_eurp)

# plot
plot(fit_eurp)

### MUDFOLD method on Plato7 data ###
data("Plato7")

## transform to binary data
## using as threshold the mean
## per row of Plato7
```

```
dat_plato <- pick(Plato7)

## fit MUDFOLD on Plato7 data ##
fit_plato <- mudfold(dat_plato)

# print
print(fit_plato)

# summary
summary(fit_plato)

# plot
plot(fit_plato)


##########################################
#### MUDFOLD method on simulated data ####
##########################################

### Data with the responses of
### n=3000 on p=20 items

simulation1 <- mudfoldsim(p=20, n=3000, pgamma1=2, pgamma2=-10, zeros=FALSE,seed = 1)
dat_sim1 <- simulation1$dat

## fit MUDFOLD on simulated data ##
fit.sim1 <- mudfold(dat_sim1)

# print
fit.sim1

# summary
summary(fit.sim1)

# plot
plot(fit.sim1)

### Data with the responses of
### n=3000 on p=26 items

simulation2 <- mudfoldsim(p=26, n=3000, gamma1=2, gamma2=-10, zeros=FALSE,seed = 1)
dat_sim2 <- simulation2$dat

## fit MUDFOLD on simulated data ##
fit.sim2 <- mudfold(dat_sim2)

# print
fit.sim2

# summary
summary(fit.sim2)
```

```
# plot
plot(fit.sim2)

## End(Not run)
```

---

mudfoldsim   *Function for constructing artificial item response data generated under a proximity response process.*

Description

mudfoldsim function simulates unfolding data following a unimodal parametric function with flexible set up. User can control the number of respondents, the number of items and fixed parameters of the Item Response Function (IRF) under which the responses are generated. Moreover, the user of the **mudfold** package can allow (or not) individuals that are endorsing no items.

Usage

```
mudfoldsim(p=p,n=n,gamma1=5,gamma2=-10,zeros=FALSE,parameters="normal",seed=NULL)
```

Arguments

p	: This argument specifies the number of items (stimuli).
n	: Argument which allows the user to specify the number of respondents in the simulated data.
gamma1	: Parameter which is used in the IRF under which the data is generated. Default value is 5.
gamma2	: Parameter which is used in the IRF under which the data is generated. Default value is -10.
zeros	: Logical argument. If zeros=FALSE (default), only individuals who endorse at least one item are allowed. Else, if zeros=TRUE individuals with no response are allowed.
parameters	: A character string that controls the distribution of the person parameters. If parameters="normal" (default), individual parameters are drawn from a standard normal distribution. If parameters="uniform", the person parameters are uniformly drawn between the minimum and the maximum item parameters respectively.
seed	: An integer to be used in the set.seed function. If seed=NULL (default), then the seed is not set.

Details

For simulating the response of an individual i with scale parameter θ_i to an item j with scale parameter β_j we use the function $P(X_j = 1 \mid \theta_i, \beta_j) = \frac{1}{1+e^{-\gamma_1-\gamma_2(\theta_i-\beta_j)^2}}$. The parameters θ_i, β_j are sampled from a standard normal distribution, i.e., $\theta \sim \mathcal{N}(0,1)$, and $\beta \sim \mathcal{N}(0,1)$.

Value

a list with 11 components.

- `obs_ord` : A character vector with the items in the simulated order.
- `true_ord` : A character vector with the items in the true order in which they constitute an unfolding scale.
- `items` : An integer corresponding to the number of the simulated items.
- `sample` : An integer corresponding to the number of the simulated respondents.
- `gamma1` : A value that corresponds to the parameter γ_1 of the IRF.
- `gamma2` : A value that corresponds to the parameter γ_2 of the IRF.
- `seed` : An integer that corresponds to the seed number that is going to be used in the `set.seed` function.
- `dat` : data frame containing the binary responses of `n` subjects on `K` items under a parametric Item Response Function.
- `probs` : A matrix containing the probabilities of positive response from `n` subjects on `K` items under a parametric Item Response Function.
- `item.patameters` : The simulated item parameters that have been used for sampling the data.
- `subject.parameters` : The simulated subject parameters that have been used for sampling the data.

Author(s)

Spyros E. Balafas (auth.), Wim P. Krijnen (auth.), Wendy J. Post (contr.), Ernst C. Wit (auth.)

Maintainer: Spyros E. Balafas (<s.balafas@rug.nl>)

References

W.H. Van Schuur.(1984). *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. CT Press.

W.J. Post. (1992). *Non parametric Unfolding Models: A Latent Structure Approach*. M & T series. DSWO Press.

W.J. Post. and T.AB. Snijders. (1993).Non parametric unfolding models for dichotomous data. Methodika.

Examples

```
## Not run:
## Simulate 5 different scenarios

n.seed <- 10

sim1 <- mudfoldsim(p=6, n=100, gamma1=5, gamma2=-10, zeros=FALSE,seed=n.seed)
sim2 <- mudfoldsim(p=10,n=1000,gamma1=10,gamma2=-100,zeros=FALSE,seed=n.seed)
sim3 <- mudfoldsim(p=15,n=2000,gamma1=50,gamma2=-100,zeros=FALSE,seed=n.seed)
sim4 <- mudfoldsim(p=30,n=2000,gamma1=50,gamma2=-100,zeros=FALSE,seed=n.seed)
```

```
sim5 <- mudfoldsim(p=50,n=2000,gamma1=50,gamma2=-100,zeros=FALSE,seed=n.seed)


dat1 <- sim1$dat
dat2 <- sim2$dat
dat3 <- sim3$dat
dat4 <- sim4$dat
dat5 <- sim5$dat

fit1 <- mudfold(dat1)
fit1
fit2 <- mudfold(dat2)
fit2
fit3 <- mudfold(dat3)
fit3
fit4 <- mudfold(dat4)
fit4
fit5 <- mudfold(dat5)
fit5

## End(Not run)
```

---

pick *Transform items to preference binary data.*

Description

Function pick can be used to transform quantitative or ordinal type of variables, into binary form (i.e., 0,1). When byItem=FALSE, then the underlying idea is that the individual selects those items with the higher preference. This is done through user provided cut-off values, or by assuming a *pick* k *out of* N response process, where, each continuous response vector takes a 1 at its k higher values. Binarization can be performed row-wise (default) or column-wise.

Usage

```
pick(x,k=NULL,cutoff=NULL,byItem=FALSE)
```

Arguments

x : A matrix or data frame containing the continuous or discrete responses of nrow(x) persons/judges to ncol(x) items. Missing values in x are not allowed.

k : An integer ($1 \leq$ k $\leq$ ncol(x)) that restricts the number of items a person can pick (default k=NULL). This argument, is used if one wants to transform the data into *pick* k *out of* N form. If k is provided by the user, cutoff should be NULL and vice verca. By default, this process is applied to the matrix x rowise. The user can restrict the number

cutoff :The value(s) that will be used as thresholds. The length of this argument should be equal to 1 (the same threshold for all rows (or columns) of x) or equal to K where K=nrow(x) or K=ncol(x) when byItem=TRUE.

`byItem` : logical argument. If byItem=TRUE, the binarization is performed columnwise. In the default byItem=FALSE, the function determines the ones rowise.

Details

Binary transformation of continuous or discrete variables with $\rho \geq 3$ number of levels. Two different methods are available for the transformation.

The first method uses the argument `k` in the `pick` function, and assumes a *pick* `k` *out of* `N` response process. Such type of response processes are met in surveys and questionnaires, in which respondents are asked to pick exactly the `k` most preferred items. The value for `k` is an integer between 1 and `ncol(x)`. By choosing an integer for `k`, this function "picks" the `k` higher values in each row (if `byItem=FALSE`) of `x`. The `k` higher values in each row become 1 and the rest `ncol(x)-k` elements are set to 0. Obviously, if `k=ncol(x)`, then the resulting matrix will only consists of 1's and no 0's.

The second method is based on thresholding in order to binarize the data. For this method, the user should provide threshold(s) with the parameter `cutoff` in the `pick` function (default `cutoff=NULL`). If one value is provided in the `cutoff` parameter, i.e., `cutoff=`α, then α is used as threshold in each row i (if `byItem=FALSE`) of the data matrix `x` such that, any value greater than or equal to `cutoff` in row i becomes 1 and 0 else. Additionally, the user can provide row (or column) specific cut off values, i.e., `cutoff=`α with $\alpha = (\alpha_1, ..., \alpha_K)$ where α_i is the cut-off value for the row or column i. In this case, if $x_{ij} \geq \alpha_i$ then $x_{ij} = 1$ and $x_{ij} = 0$ else.

The two methods cannot be used simultaneously. Only one of the parameters `k` and `cutoff` can be different than `NULL` each time. If both parameters are equal `NULL` (default), then a row specific cut off is determined automatically for each row i of `x`, such that, $\alpha_i = \bar{x_i}$. The binarization is performed by row of `x`, except the case, `byItem=TRUE`.

When the argument `k` is used, it can be the case that more than `k` values can be picked (i.e., ties). In this case, the choice on which item will be picked is being made after we add a small amount of noise in each observation of row or column i. This is done with the function `jitter`.

Value

Binary valued (i.e., 0-1) data with the same dimensions as the input.

Author(s)

Spyros E. Balafas (auth.), Wim P. Krijnen (auth.), Wendy J. Post (contr.), Ernst C. Wit (auth.)

Maintainer: Spyros E. Balafas (`<s.balafas@rug.nl>`)

Examples

```
## Not run:
### simulate some data with 3 discrete variables with three levels
### and 1 variable with 4 levels
d1 <- cbind(sample(1:3,20,replace = TRUE),
            sample(1:3,20,replace = TRUE,prob = c(0.3,0.3,0.4)),
            sample(1:3,20,replace = TRUE,prob = c(0.2,0.4,0.4)),
            sample(1:4,20,replace = TRUE,prob = c(.1,.3,.4,.2)))


### apply pick on d1 ###
```

```
# binarize at the mean of
# each row and column
d1_rowmean <- pick(d1)
d1_colmean <- pick(d1,byItem = TRUE)

# binarize at the cutoff=2
d1_cut <- pick(d1,cutoff = 2,byItem = TRUE)

# binarize at different cutoffs (per row)
# for example at the median of each row
med_cuts <- apply(d1,1,median)
d1_cuts <- pick(d1,cutoff = med_cuts)

# binarize at different cutoffs (per column)
# for example at the median of each column
med_cuts_col <- apply(d1,2,median)
d1_cuts_col <- pick(d1,cutoff = med_cuts_col,byItem = TRUE)


# binarize at the k=2 higher values
# per row and column
d1_krow <- pick(d1,k = 2)
d1_kcol <- pick(d1,k = 2,byItem = TRUE)

## End(Not run)
```

Plato7 *Plato's Seven Works*

Description

This dataset contains statistical information about Plato's seven works. The underlying problem to this dataset is the fact that the chronological order of Plato's works is unknown. Scholars only know that Republic was his first work, and Laws his last work. For each work, Cox and Brandwood (1959) extracted the last five syllables of each sentence. Each syllable is classified as long or short which gives 32 types. Consequently, we obtain a percentage distribution across the 32 scenarios for each of the seven works. The dataset has been borrowed from the package **smacof** (De Leeuw and Mair, 2009).

Usage

```
data(Plato7)
```

Format

Data frame containing syllable percentages of Plato's 7 works.

References

Cox, D. R. & Brandwood, L. (1959). On a discriminatory problem connected with the work of Plato. Journal of the Royal Statistical Society (Series B), 21, 195-200.

De Leeuw, J.& Mair, P. (2009). Multidimensional Scaling Using Majorization: SMACOF in R. Journal of Statistical Software, 31(3), 1-30. URL http://www.jstatsoft.org/v31/i03/.

Examples

```
## Not run:
data(Plato7)
str(Plato7)

## End(Not run)
```

`plot.mdf` — `plot` *function for* `mdf` *class objects.*

Description

Generic function for plotting S3 class `mdf` objects. This function, is plotting the rows of the conditional adjacency matrix (CAM) which are nonparametric estimates of the item response functions. The plot is produced using the `ggplot` function from the package **ggplot2**.

Usage

```
## S3 method for class 'mdf'
plot(x,select=NULL,plot.type="IRF",...)
```

Arguments

`x`	Object of class `mdf`
`select`	: in this argument we can provide a subset of items that we wish them to be explicitly plotted. If `select=NULL` the estimated IRF for each item in the scale is plotted.
`plot.type`	: Determines the type of plot that is returned. By default, `plot.type="IRF"`, which returns the estimated IRFs for the MUDFOLD item order. The user can set `plot.type="scale"` in order to get plotted the unidimensional MUDFOLD scale.
`...`	Other arguments passed on to `ggplot` plotting method.

Details

The `plot` method is used to obtain a graphical representation of the estimated item response functions. As estimates of the IRFs are considered the rows of the CAM. For interpolating the missing diagonal elements of the CAM, we make use of the `na.approx` function from the package **zoo**.

Author(s)

Spyros E. Balafas (auth.), Wim P. Krijnen (auth.), Wendy J. Post (contr.), Ernst C. Wit (auth.)

Maintainer: Spyros E. Balafas (<s.balafas@rug.nl>)

References

W.H. Van Schuur.(1984). *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. CT Press.

W.J. Post. (1992). *Nonparametric Unfolding Models: A Latent Structure Approach*. M & T series. DSWO Press.

W.J. Post and T.AB. Snijders. (1993).Nonparametric unfolding models for dichotomous data. Methodika.

A. Zeileis and G. Grothendieck. (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27. doi:10.18637/jss.v014.i06

H. Wickham. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

H. Wickham. (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.

Examples

```
## Not run:
data(ANDRICH)
fit <- mudfold(ANDRICH)
plot(fit)
plot(fit,select="DONTBELIEV")

## End(Not run)
```

---

print.mdf — `print` *method for* `mdf` *class objects.*

Description

S3 generic function for printing `mdf` class objects.

Usage

```
## S3 method for class 'mdf'
print(x, Item.info= FALSE, Diagnostics= FALSE,...)
```

Arguments

`x`	Object of class `mdf`
`Item.info`	logical argument. If `Item.info=TRUE` then additional information for each item in the scale is printed.
`Diagnostics`	logical argument. If `Diagnostics= TRUE` then the diagnostic matrices for the item order are printed.
`...`	further arguments passed on to the `print` method.

Author(s)

Spyros E. Balafas (auth.), Wim P. Krijnen (auth.), Wendy J. Post (contr.), Ernst C. Wit (auth.)

Maintainer: Spyros E. Balafas (<s.balafas@rug.nl>)

References

W.H. Van Schuur.(1984). *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. CT Press.

W.J. Post. (1992). *Nonparametric Unfolding Models: A Latent Structure Approach*. M & T series. DSWO Press.

W.J. Post. and T.AB. Snijders (1993). *Nonparametric unfolding models for dichotomous data*. Methodika.

Examples

```
## Not run:
data(ANDRICH)
fit <- mudfold(ANDRICH)
fit
print(fit,Item.info= TRUE, Diagnostics= TRUE)

## End(Not run)
```

`summary.mdf` — `summary` *method for S3 class* `mdf` *objects.*

Description

Generic function that is used in order to summarize information from `mdf` class objects.

Usage

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

Arguments

`object`	Object of class `mdf` resulted from the function `mudfold` or `as.mudfold`.
`...`	Other arguments passed on to method `summary`. Not currently used.

Details

A summary of the MUDFOLD scale that has been calculated with the `mudfold` function.

Value

returns a data frame with parameters concerning the fit of each item in the scale as well as for the quality of the scale as a whole. The columns of the resulting data frame are the following.

`index`	: `m` selected items in the estimated MUDFOLD rank order. In the cell `total`, information for the total scale is given wherever this is available.
`samp.size`	: The sample size of the data.
`freq`	: observed frequency of positive responses.
`prop`	: observed proportion of positive responses.
`std.err`	: standard error for the proportion of positive responses.
`Obs.err`	: observed response errors.
`Exp.err`	: expected response errors.
`Iso`	: iso statistic.
`Scalab.H`	: scalability coefficients.

Author(s)

Spyros E. Balafas (auth.), Wim P. Krijnen (auth.), Wendy J. Post (contr.), Ernst C. Wit (auth.)

Maintainer: Spyros E. Balafas (<s.balafas@rug.nl>)

References

W.H. Van Schuur.(1984). *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. CT Press.

W.J. Post. (1992). *Nonparametric Unfolding Models: A Latent Structure Approach*. M & T series. DSWO Press.

W.J. Post. and T.AB. Snijders (1993). *Nonparametric unfolding models for dichotomous data*. Methodika.

Examples

```
## Not run:
data(ANDRICH)
fit <- mudfold(ANDRICH)
summary(fit)

## End(Not run)
```

Index

∗Topic **datasets**
  ANDRICH, 3
  EURPAR2, 7
  Plato7, 17
∗Topic **methods**
  plot.mdf, 18
  print.mdf, 19
  summary.mdf, 20
∗Topic **package**
  mudfold-package, 2

ANDRICH, 3
as.mudfold, 4

EURPAR2, 7

mudfold, *2*, *6*, 8, *21*
mudfold-package, 2
mudfoldsim, 13

pick, 15
Plato7, 17
plot.mdf, 18
print.mdf, 19

summary.mdf, 20