Package 'TestGardener'

November 24, 2022

Version 3.1.0

Date 2022-11-22

Type Package

Title Optimal Analysis of Test and Rating Scale Data

Maintainer James Ramsay <james.ramsay@mcgill.ca>

Depends R (>= 3.5), fda, ggplot2, rgl

Description Develop, evaluate, and score multiple choice examinations, psychological scales, questionnaires, and similar types of data involving sequences of choices among one or more sets of answers.
This version of the package should be considered as brand new. Almost all of the functions have been changed, including their argument list.
See the file NEWS.Rd in the Inst folder for more information.
Using the package does not require any formal statistical knowledge beyond what would be provided by a first course in statistics in a social science department. There the user would encounter the concept of probability and how it is used to model data and make decisions, and would become familiar with basic mathematical and statistical notation.
Most of the output is in graphical form.

License GPL (>= 2)

LazyData true

Imports dplyr, ggpubr, stringr, tidyr, pracma, utf8, knitr, rmarkdown

VignetteBuilder knitr

BuildVignettes yes

Language en-US

NeedsCompilation no

Author James Ramsay [aut, cre],
Juan Li [ctb],
Marie Wiberg [ctb],
Joakim Wallmark [ctb],
Spencer Graves [ctb]

Repository CRAN

Date/Publication 2022-11-24 14:50:02 UTC

R topics documented:

Analyze *Analyze test or rating scale data defined in* dataList.

Description

The test or rating scale data have already been processed by function make.dataList or or other code to produce the list object dataList. The user defines a list vector ParameterList which stores results from a set of cycles of estimating surprisal curves followed by estimating optimal score index values for each examinee or respondent. These score index values are within the interval [0,100]. The number of analysis cycles is the length of the parList list vector.

Usage

```
Analyze(theta, thetaQnt, dataList, ncycle=10, itdisp=FALSE, verbose=FALSE)
```

Arguments

theta	A vector of N score index values for the examinees or respondents. These values are in the percent interval [0,100].
thetaQnt	A vector of length 2*nbin + 1 where nbin is the number of bins containing score index values. The vector begins with the lower boundary 0 and ends with the upper boundary 100. In between it alternates between the bin center value and the boundary separating the next bin.
dataList	A list that contains the objects needed to analyse the test or rating scale with the following fields:

U: A matrix of response data with N rows and n columns where N is the number of examinees or respondents and n is the number of items. Entries in the matrices are the indices of the options chosen. Column i of U is expected to contain only the integers 1,...,noption.

optList: A list vector containing the numerical score values assigned to the options for this question.

key: If the data are from a test of the multiple choices type where the right answer is scored 1 and the wrong answers 0, this is a numeric vector of length n containing the indices the right answers. Otherwise, it is NULL.

WfdPar: An fdPar object for the defining the surprisal curves.

noption: A numeric vector of length n containing the numbers of options for each item.

nbin: The number of bins for binning the data.

scrrng: A vector of length 2 containing the limits of observed sum scores.

scrfine: A fine mesh of test score values for plotting.

scrvec: A vector of length N containing the examinee or respondent sum scores.

itemvec: A vector of length n containing the question or item sum scores.

percntrnk: A vector length N containing the sum score percentile ranks.

thetaQnt: A numeric vector of length 2*nbin + 1 containing the bin boundaries alternating with the bin centers. These are initially defined as seq(0,100,len=2*nbin+1).

Wdim: The total dimension of the surprisal scores.

PcntMarkers: The marker percentages for plotting: 5, 25, 50, 75 and 95.

`ncycle` The number of cycles executed by `Analyze`.

`itdisp` If TRUE, the progress of the iterations within each cycle for estimating theta are reported.

`verbose` If TRUE, the stages of analysis within each cycle for estimating theta are reported.

Details

The cycling process is described in detail in the references, and displayed in R code in the vignette `SweSATQuantitativeAnalysis`.

Value

The list vector `parList` where each member is a named list object containing the results of an analysis cycle. These results are:

theta: The optimal estimates of the score index values for the examinees/respondents. This is a vector of length `N`.

thetaQnt: A vector of length 2*nbin+1 containing bin boundaries alternating with bin edges.

WfdList: A list vector containing results from the estimation of surprisal curves. The list vector is of length `n`, the number of questions or items in the test of rating scale. For details concerning these results, see function `Wbinsmth()`.

logdensfd: A functional data object defining the estimate of the log of the probability density function for the distribution of the score index values.

C: The normalizing value for probability density functions. A density value is computed by dividing the exponential of the log density value by this constant.

densfine: The value of the probability density function over a fine mesh of 101 equally spaced score index values.

denscdf: The values over a fine mesh of the cumulative probability distribution function. These values start at 0 and end with 1 and are increasing. Ties are often found at the upper boundary, so that using these values for interpolation purposes may require using the vector `unique(denscdf)`.

binctr: A vector of length `nbin` containing the bin centers within the interval [0,100].

bdry: A vector of length `nbin+1` containing the bin boundaries.

freq: A vector of length `nbin` containing the number of score index values in the bins. An score index value is within a bin if it is less than or equal to the upper boundary and greater than the lower boundary. The first boundary also contains zero values.

Hval: A vector of length N containing the values of the negative log likelihood fitting criterion.

DHval: A vector of length N containing the values of the first derivative of the negative log likelihood fitting criterion.

D2Hval: A vector of length N containing the values of the second derivative of the negative log likelihood fitting criterion.

active: A vector of length N of the activity status of the values of theta. If convergence was not achieved, the value is TRUE, otherwise FALSE.

arclength: The length of the space curve defined by the surprisal curves.

alfine: A vector of length 101 of arclengths corresponding to equally spaced values of theta.

theta_al: A vector of length N of arclengths corresponding to estimated values of theta.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

make.dataList, dataSimulation, Power.plot, scoreDensity, Sensitivity.plot, testscore, theta.distn, theta2arclen, thetafun, Wbinsmth, ICC.plot, Wpca.plot

---

`ArcLength.plot` *Plot the score index* `theta` *as a function of arc length.*

Description

Arc length is the distance along the space curved traced out as score index `theta` increases from 0 to 100. It is measured in bits and is remains unchanged if the score index continuum is modified.

Usage

```
ArcLength.plot(arclength, arclengthvec, titlestr=NULL)
```

Arguments

`arclength`	This is a positive real number indicating the total length of the space curve. It is expressed in terms of numbers of bits.
`arclengthvec`	A vector of length 101 containing equally-spaced arc-length distances along the test information curve.
`titlestr`	A string for the title of the data.

Value

A `gg` or `ggplot` object defining the plot of arclength along the test information curve as a function of the score index `theta`. This is displayed by the `print` command. The plot is automatically displayed as a side value even if no return object is specified in the calling statement.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

theta2arclen

Examples

```
#  Example 1.  Display the arc length curve for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
indfine       <- seq(0,100,len=101)
WfdList       <- Quantshort_parList$WfdList
theta         <- Quantshort_parList$theta
Qvec          <- Quantshort_parList$Qvec
binctr        <- Quantshort_parList$binctr
arclenList    <- theta2arclen(theta, Qvec, WfdList, binctr)
names(arclenList)
arclength    <- arclenList$arclength
arclengthvec <- arclenList$arclengthvec
oldpar <- par(no.readonly=TRUE)
on.exit(oldpar)
ArcLength.plot(arclength, arclengthvec)
#  Example 2.  Display the arc length curve for the
#  Symptom Distress Scale with 13 items and 473 respondents
indfine       <- seq(0,100,len=101)
WfdList       <- SDS_parList$WfdList
theta         <- SDS_parList$theta
Qvec          <- SDS_parList$Qvec
binctr        <- SDS_parList$binctr
arclenList    <- theta2arclen(theta, Qvec, WfdList, binctr)
arclength    <- arclenList$arclength
arclengthvec <- arclenList$arclengthvec
ArcLength.plot(arclength, arclengthvec)
```

`dataSimulation` *Simulation Based Estimates of Root-mean-squared-err of Theta Estimates*

Description

Estimate sum score,s score index values theta and test information values bias and mean squared errors using simulated data.

Usage

```
dataSimulation(dataList, parList, theta.pop = seq(0, 100, len = 101),
               nsample = 1000)
```

Arguments

dataList	The list object set up by function make_dataList.
parList	The list object containing objects compuated by function Analyze.
theta.pop	A vector containing true values of theta to be estimated using simulated data.
nsample	The number of simulated samples.

Value

A named list object containing objects produced from analyzing the simulations, one set for each simulation:

sumscr: Sum score estimates

theta: Score index estimates

mu: Expected sum score estimates

al: Total arc length estimates

thepop: True or population score index values

mupop: Expected sum score population values

alpop: Total test length population values

n: Number of items

ntheta: Number of theta values

indfine: Fine mesh over score index range

Qvec: Five marker percentages: 5, 25, 50, 75 and 95

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

scorePerformance

---

density_plot *Plot the probability density function for a set of test scores*

Description

Plots the probability density function of a set of score values that are not at the score boundaries as a smooth curve, and also plots the proportions of score values at both boundaries as points. The score values are typically either the values of the score index values theta or the arclength or information score values.

Usage

```
density_plot(scrvec, scrrng, Qvec, xlabstr=NULL, titlestr=NULL,
                          scrnbasis=15, nfine=101)
```

Arguments

scrvec	A vector of N score values
scrrng	A vector of length 2 containing boundary values
Qvec	A vector of length 5 containing the score values corresponding to the marker percentages 5, 25, 50, 75 and 95.
xlabstr	Label for abscissa
titlestr	Label for plot
scrnbasis	The number of spline basis functions used for representing the smooth density function
nfine	Number of plotting points

Value

A plot object.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

scoreDensity

Examples

```
#  Example 1.  Display probability density curve for the
#  score index values for the short SweSAT multiple choice test with
#  24 items and 1000 examinees
theta   <- Quantshort_parList$theta
Qvec    <- Quantshort_parList$Qvec
#  plot the density for the score indices within interval c(0,100)
oldpar <- par(no.readonly=TRUE)
on.exit(oldpar)
density_plot(theta, c(0,100), Qvec, xlabstr="Score index",
             titlestr="SweSAT 13B Theta Density",
             scrnbasis=11, nfine=101)
#  Example 2.  Display probability density curve for the
#  arc length or information values
theta_al  <- Quantshort_infoList$theta_al
Qvec_al   <- Quantshort_infoList$Qvec_al
arclength <- Quantshort_infoList$arclength
#  plot the density for the score indices within interval c(0,arclength)
density_plot(theta_al, c(0,arclength), Qvec_al, xlabstr="Score index",
             titlestr="SweSAT 13B Theta Density",
             scrnbasis=11, nfine=101)
```

DHfun *Compute the first and second derivatives of the negative log likelihoods*

Description

DHfun computes the first and second derivatives of the negative log likelihoods for a set of examinees. Items can be either binary or multi-option. The analysis is within the closed interval [0,100].

Usage

```
DHfun(theta, WfdList, Umat)
```

Arguments

theta	Initial values for score indices in [0,n]/[0,100]. Vector of size N.
WfdList	A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of WfdList is a named list containing information computed during the analysis. These named lists contain these objects: **Wfd:** A functional data object containing the M surprisal curves for a question. **M:** The number of options.

Pbin: A matrix containing proportions at each bin.

Wbin: A matrix containing surprisal values at each bin.

Pmatfine: A matrix of probabilities over a fine mesh.

Wmatfine: A matrix of surprisal values over a fine mesh.

DWmatfine: A matrix of the values of the first derivative of surprisal curves over fine mesh.

D2Wmatfine: A matrix of the values of the second derivative of surprisal curves over fine mesh.

`Umat` An `N` by `n` matrix of responses. If `N` = 1, it can be a vector of length `n`.

Value

A named list for results `DH` and `D2H`:

DH: First derivatives of the negative log likelihood values, vector of size N

D2H: Second derivatives of the negative log likelihood values, vector of size N

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

`make.dataList`, `Hfun`, `Hfuns.plot`

Examples

```
#  Example 1:
#  Compute the first and second derivative values of the objective function for
#  locating each examinee for the 24-item short form of the SweSAT quantitative
#  test on thepercentile score index continuum.
WfdList <- Quantshort_parList$WfdList
theta   <- Quantshort_parList$theta
U       <- Quantshort_dataList$U
DHfunResult <- DHfun(theta, WfdList, U)
DHval  <- DHfunResult$DH
D2Hval <- DHfunResult$D2H
print(paste("Mean               of objective gradient =",round(mean(DHval),4)))
print(paste("Standard deviation of objective gradient =",round(sqrt(var(DHval)),4)))
print(paste("Mean               of objective Hessian =",round(mean(D2Hval),4)))
print(paste("Standard deviation of objective Hessian =",round(sqrt(var(D2Hval)),6)))
```

```
#  Example 2:
#  Compute the first and second derivative values of the objective function for
#  locating each examinee for the 13-item Symptom Distress scale
#  on the percentile score index continuum.
#  Proceed as above changing "Quant" for "SDS".
```

`entropies` *Entropy measures of inter-item dependency*

Description

Entropy I_1 is a scalar measure of how much informationn is required to predict the outcome of a choice number 1 exactly, and consequently is a measure of item effectiveness suitable for multiple choice tests and rating scales. Joint entropy $J_1,2$ is a scalar measure of the cross-product of multinomial vectors 1 and 2. Mutual entropy $I_1,2 = I_1 + I_2 - J_1,2$ is a measure of the co-dependency of items 1 and 2, and thus the analogue of the negative log of a squared correlation R^2. this function computes all four types of entropies for two specificed items.

Usage

```
entropies(theta, m, n, U, noption)
```

Arguments

`theta`	A vector of length N containing score index values for each test taker.
`m`	The index of the first choice.
`n`	The index of the second choice.
`U`	The data matrix containing the indices of choisen options for each test taker.
`noption`	A vector containing the number of options for all items.

Value

A named list object containing objects produced from analyzing the simulations, one set for each simulation:

I_m: The entropy of item m.

I_n: The entropy of item n.

J_nm: The joint entropy of items m and n.

I_nm: The mutual entropy of items m and n.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

Entropy.plot

Examples

```
#  Load needed objects
U       <- Quantshort_dataList$U
theta   <- Quantshort_parList$theta
noption <- matrix(5,24,1)
#  compute mutual entropies for all pairs of the first 6 items
Mvec    <- 1:6
Mlen    <- length(Mvec)
Hmutual <- matrix(0,Mlen,Mlen)
for (i1 in 1:Mlen) {
  for (i2 in 1:i1) {
    Result <- entropies(theta, Mvec[i1], Mvec[i2], U, noption)
    Hmutual[i1,i2] = Result$Hmutual
    Hmutual[i2,i1] = Result$Hmutual
  }
}
print("Matrix of mutual entries (off-digagonal) and self-entropies (diagonal)")
print(round(Hmutual,3))
```

Entropy.plot *Plot item entropy curves for selected items or questions.*

Description

Item the value of the entropy curve at a point `theta` is the expected value of the surprisal curve values. Entropy is a measure of the randomness of the surprisal value, which is maximized when all the surprisal curves have the same value and has a minimum of zero if all but a single curve has probability zero. This is unattainable in the calculation, but can be arbitrarily close to this state.

Usage

```
Entropy.plot(scrfine, WfdList, Qvec, dataList, plotindex=1:n,
             plotrange=c(min(scrfine),max(scrfine)), height=1.0, value=0,
             ttlsz=NULL, axisttl=NULL, axistxt=NULL)
```

Arguments

scrfine	A vector of length nfine (usually 101) containing equally spaced points spanning the plotrange. Used for plotting.
WfdList	A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of WfdList is a named list containing information computed during the analysis. These named lists contain these objects: **Wfd:** A functional data object containing the M surprisal curves for a question. **M:** The number of options. **Pbin:** A matrix containing proportions at each bin. **Wbin:** A matrix containing surprisal values at each bin. **Pmatfine:** A matrix of probabilities over a fine mesh. **Wmatfine:** A matrix of surprisal values over a fine mesh. **DWmatfine:** A matrix of the values of the first derivative of surprisal curves over fine mesh. **D2Wmatfine:** A matrix of the values of the second derivative of surprisal curves over fine mesh.
Qvec	The five marker percentile values.
dataList	A list vector containing objects essential to an analysis.
plotindex	A set of integers specifying the numbers of the items or questions to be displayed.
plotrange	A vector of length 2 containing the plot boundaries within or over the score index interval c(0,100).
height	A positive real number defining the upper limit on the ordinate for the plots.
value	Number required by ggplot2. Defaults to 0.
ttlsz	Title font size.
axisttl	Axis title font size.
axistxt	Axis text(tick label) font size.

Details

An entropy curve for each question indexed in the index argument. A request for a keystroke is made for each question. The answer to question strongly defines the optimal position of an estimated score index value where the curve is high value. Values of entropy curves typically range over [0,1].

Value

The plots of the entropy curves specified in plotindex are produced as a side effect. If saveplot is TRUE, the plots of item entropy curves specified in plotindex are bundled into a single postscript or .pdf file and the file name is defined by paste(dataList$titlestr,i,'-entropy.pdf',sep="").
The file is then output as a returned value.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

Sensitivity.plot, Power.plot, Hfuns.plot, ICC.plot

Examples

```
#  Example 1.  Display the item entropy curves for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
#  plot the entropy curve for the first item
WfdList <- Quantshort_parList$WfdList
Qvec    <- Quantshort_parList$Qvec
scrfine <- seq(0,100,len=101)
oldpar <- par(no.readonly=TRUE)
Entropy.plot(scrfine, WfdList, Qvec, Quantshort_dataList, plotindex=1)
par(oldpar)
#  Example 2.  Display the item entropy curves for the
#  Symptom Distress Scale with 13 items and 473 respondents
#  Proceed as above changing "Quant" to "SDS".
```

Hcurve *Construct grid of 101 values of the fitting function*

Description

A fast grid of values of the fitting function or one of its first two derivatives is constructed for use in function thetasearch.

Usage

```
Hcurve(WfdList, U, nderiv=0)
```

Arguments

WfdList	A list vector containing specifications of surprisal curves for each item.
U	A N by n matrix containing indices of chosen items for each test taker.
nderiv	Integer 0, 1 or 2 to indicate which level of derivative to use.

Value

A vector of length 101 containing grid values of a derivative of the fitting function

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

thetasearch

Examples

```
#  Compute a grid of values of the objective function for locating each
#  examinee or respondent for the 24-item short form of the SweSAT quantitative
# test on the percentile score index continuum [0,100].
WfdList <- Quantshort_parList$WfdList
theta   <- Quantshort_parList$theta
U       <- Quantshort_dataList$U
n       <- ncol(U)
#  Fitting function for the first examinee
j <- 1
Umat    <- matrix(U[j,],1,n)
Hcurve1   <- Hcurve(WfdList, Umat, 0)
#  First derivative of the fitting function for the first examinee
DHcurve1  <- Hcurve(WfdList, Umat, 1)
#  Second derivative of the fitting function for the first examinee
D2Hcurve1 <- Hcurve(WfdList, Umat, 2)
oldpar <- par(no.readonly=TRUE)
par(mfrow=c(3,1))
indfine <- seq(0,100,len=101)
plot(indfine, Hcurve1,   type="l", xlab="", ylab="Fitting curve",
     main="Examinee 1")
plot(indfine, DHcurve1,  type="l", xlab="", ylab="First derivative")
points(theta[1], 0, pch="o")
abline(0,0,lty=2)
plot(indfine, D2Hcurve1, type="l", xlab="Score index", ylab="Second derivative")
abline(0,0,lty=2)
points(theta[1], 0, pch="o")
par(oldpar)
```

Hfun *Compute the negative log likelihoods associated with a vector of score index values.*

Description

Hfun computes the negative log likelihoods for a set of examinees, each at a single value theta.

Usage

```
Hfun(theta, WfdList, Umat)
```

Arguments

theta	A vector of size N containing values for score indices in the interval [0,100].
WfdList	A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of WfdList is a named list containing information computed during the analysis. These named lists contain these objects: **Wfd:** A functional data object containing the M surprisal curves for a question. **M:** The number of options. **Pbin:** A matrix containing proportions at each bin. **Wbin:** A matrix containing surprisal values at each bin. **Pmatfine:** A matrix of probabilities over a fine mesh. **Wmatfine:** A matrix of surprisal values over a fine mesh. **DWmatfine:** A matrix of the values of the first derivative of surprisal curves over fine mesh. **D2Wmatfine:** A matrix of the values of the second derivative of surprisal curves over fine mesh.
Umat	An N by n matrix of responses or, for a single examinee, a vector of length n.

Value

A vector of length N of negative log likelihood values.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

make.dataList, Hfun, Hfuns.plot

Examples

```
#  Example 1:  Compute the values of the objective function for locating each
#  examinee or respondent for the 24-item short form of the SweSAT quantitative
# test on the percentile score index continuum [0,100].
WfdList <- Quantshort_parList$WfdList
theta   <- Quantshort_parList$theta
U       <- Quantshort_dataList$U
Hval    <- Hfun(theta, WfdList, U)
print(paste("Mean               of objective function =",round(mean(Hval),2)))
print(paste("Standard deviation of objective function =",round(sqrt(var(Hval)),2)))
#  Example 2:  Compute the values of the objective function for locating each
#  examinee or respondent for the 13-item Symptom Distress Scale
# on the percentile score index continuum [0,100].
#  Proceed as above changing "Quant" for "SDS".
```

Hfuns.plot *Plot a selection of fit criterion H functions and their first two derivatives.*

Description

These plots indicate whether an appropriate minimum of the fitting criterion was found. The value of theta should be at the function minimum, the first derivative be close to zero there, and the second derivative should be positive. If these conditions are not met, it may be worthwhile to use function thetafun initialized with an approximate minimum value of score index theta to re-estimate the value of theta.

Usage

```
Hfuns.plot(theta, WfdList, U, plotindex=1)
```

Arguments

theta	The entire vector of estimated values of theta.
WfdList	The list vector of length n containing the estimated surprisal curves.
U	The entire N by n matrix of choice indices.
plotindex	A subset of the integers 1:N.

Details

The curves are displayed in three vertically organized panels along with values of theta and the values and first two derivative values of the fit criterion. If more than one index value is used, a press of the Enter or Return key moves to the next index value.

Value

A list vector is returned which is of the length of argument `plotindex`. Each member of the vector is a `gg` or `ggplot` object for the associated `plotindex` `value`. Each plot can be displayed using the `print` command. The plots of item power are produced as a side value even if no output object is specified in the call to the function.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

`thetafun`, `Hfun`, `DHfun`

Examples

```
#  Example 1.  Display fit criterion values and derivatives for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
index <- 1
theta    <- Quantshort_parList$theta
WfdList  <- Quantshort_parList$WfdList
U        <- Quantshort_dataList$U
Hval     <-  Hfun(theta[index], WfdList, U[index,])
DHResult <- DHfun(theta[index], WfdList, U[index,])
DHval    <- DHResult$DH
D2Hval   <- DHResult$D2H
Hfuns.plot(theta, WfdList, U, plotindex=index)
#  Example 2.  Compute score index values theta for the
#  Symptom Distress Scale with 13 items and 473 respondents
#  Proceed as above changing "Quant" to "SDS".
```

ICC.fit *Estimate the option probability and surprisal curves.*

Description

The surprisal curves for each item are fit to the surprisal transforms of choice probabilities for each of a set of bins of current performance values `theta`. The error sums of squares are minimized by the surprisal optimization `smooth.surp` in the `fda` package. The output is a list vector of length `n` containing the functional data objects defining the curves.

Usage

```
ICC.fit(theta, dataList, WfdList=dataList$WfdList,
                    thetaQnt=seq(0,100, len=2*nbin+1), wtvec=matrix(1,n,1),
                    iterlim=20, conv=1e-4, dbglev=0)
```

Arguments

`theta` A vector of length N containing current values of score index percentile values.

`dataList` A list that contains the objects needed to analyse the test or rating scale with the following fields:

U: A matrix of response data with N rows and n columns where N is the number of examinees or respondents and n is the number of items. Entries in the matrices are the indices of the options chosen. Column i of U is expected to contain only the integers `1,...,noption`.

optList: A list vector containing the numerical score values assigned to the options for this question.

key: If the data are from a test of the multiple choices type where the right answer is scored 1 and the wrong answers 0, this is a numeric vector of length n containing the indices the right answers. Otherwise, it is NULL.

WfdPar: An fdPar object for the defining the surprisal curves.

noption: A numeric vector of length n containing the numbers of options for each item.

nbin: The number of bins for binning the data.

scrrng: A vector of length 2 containing the limits of observed sum scores.

scrfine: A fine mesh of test score values for plotting.

scrvec: A vector of length N containing the examinee or respondent sum scores.

itemvec: A vector of length n containing the question or item sum scores.

percntrnk: A vector length N containing the sum score percentile ranks.

thetaQnt: A numeric vector of length 2*nbin + 1 containing the bin boundaries alternating with the bin centers. These are initially defined as `seq(0,100,len=2*nbin+1)`.

Wdim: The total dimension of the surprisal scores.

PcntMarkers: The marker percentages for plotting: 5, 25, 50, 75 and 95.

`WfdList` A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of `WfdList` is a named list containing information computed during the analysis. These named lists contain these objects:

Wfd: A functional data object containing the `M` surprisal curves for a question.

M: The number of options.

Pbin: A matrix containing proportions at each bin.

Wbin: A matrix containing surprisal values at each bin.

indfine: A vector of length 101 containing equally spaced plotting points.

Pmatfine: A matrix of probabilities over a fine mesh.

Wmatfine: A matrix of surprisal values over a fine mesh.

DWmatfine: A matrix of the values of the first derivative of surprisal curves over fine mesh.

D2Wmatfine: A matrix of the values of the second derivative of surprisal curves over fine mesh.

By default `ICC.fit` uses the `WfdList` member of `dataList`, which is generated by the initialization with `ICC.fit.init` within `make.dataList`. Usually this is what is required. However, `WfdList` may occasionally need to be entered from another source, and so included here as a separate argument.

`thetaQnt` A vector of length 2*n+1 containing the sequence of bin boundary and bin centre values.

`wtvec` A vector of length `n` of weights on observations. Defaults to all ones.

`iterlim` The maximum number of iterations used in optimizing surprisal curves. Defaults to 20.

`conv` Convergence tolerance. Defaults to 0.0001.

`dbglev` Level of output within `ICC.fit`. If 0, no output, if 1 the error sum of squares and slope on each iterations, and if 2 or higher, results for each line search iteration with function `lnsrch`.

Details

The function first bins the data in order to achieve rapid estimation of the option surprisal curves. The argument `thetaQnt` contains the sequence of bin boundaries separated by the bin centers, so that it is of length `2*nbin + 1` where `nbin` is the number of bins. These bin values are distributed over the percentile interval [0,100] so that the lowest boundary is 0 and highest 100. Prior to the call to `ICC.fit` these boundaries are computed so that the numbers of values of `theta` falling in the bins are roughly equal. It is important that the number of bins be chosen so that the bins contain at least about 25 values.

After the values of `theta` are binned, the proportions that the bins are chosen for each question and each option are computed. Proportions of zero are given NA values.

The positive proportions are then converted to surprisal values where surprisal = -log_M (proportion) where log_M is the logarithm with base M, the number of options associated with a question. Bins with zero proportions are assigned a surprisal that is appropriately large in the sense of being in the range of the larger surprisal values associated with small but positive proportions.

The next step is to fit the surprisal values for each question by a functional data object that is smooth, passes as closely as possible to an option's surprisal values, and has values consistent with being a surprisal value. The function `smooth.surp()` is used for this purpose.

Finally the curves and other results for each question are saved in object `WfdList`, a list vector of length n, and the list vector is returned.

Value

The optimized numbered list object produced by `ICC.fit`. See above for its description.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

ICC.plot, thetafun

Examples

```
#  Example 1.  Display the item probability and surprisal curves for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
#  estimate the initial surprisal curves
theta     <- Quantshort_dataList$percntrnk
thetaQnt  <- Quantshort_parList$thetaQnt
Qvec      <- Quantshort_parList$Qvec
Result    <- ICC.fit(theta, Quantshort_dataList)
WfdList   <- Result$WfdList
#  plot the curves for the first question
binctr    <- Quantshort_parList$binctr
scrfine   <- seq(0,100,len=101)
ICC.plot(scrfine, WfdList, Quantshort_dataList, Qvec, binctr,
         plotindex=1, plotrange=c(0,100))
#  estimate the final surprisal curves
theta    <- Quantshort_parList$theta
thetaQnt <- Quantshort_parList$thetaQnt
Qvec     <- Quantshort_parList$Qvec
WfdList  <- Quantshort_parList$WfdList
#  plot the curves for the first question
ICC.plot(scrfine, WfdList, Quantshort_dataList, Qvec, binctr,
         plotindex=1, plotrange=c(0,100))
#  Example 2.  Display the item probability and surprisal curves for the
# Symptom Distress Scale with 13 items and 473 respondents.
# Proceed as above changing "Quant" to "SDS"
```

ICC.plot *Plot probability and surprisal curves for a selection of test or scale items.*

Description

ICC.plots plots each item in argument `plotindex` in turn after function`Wbinsmth()` has used spline smoothing to estimate item and option characteristic curves.

Usage

```
ICC.plot(scrfine, WfdList, dataList, Qvec, binctr, plotType = "P",
         plotindex=1:n, plotrange=c(min(scrfine),max(scrfine)),
         shaderange = NULL, Wrng=c(0,5), DWrng=c(-0.2, 0.2),
         data_point = FALSE, ci = FALSE,
         titlestr = NULL, autoplot = FALSE, ttlsz = NULL,
         axisttl = NULL, axistxt = NULL, lgdlab = NULL)
```

Arguments

`scrfine` A vector of 101 plotting points.

`WfdList` A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of `WfdList` is a named list containing information computed during the analysis. These named lists contain these objects:

scrfine: A vector of length 101 containing plotting points.

Wfd: A functional data object containing the `M` surprisal curves for a question.

M: The number of options.

Pbin: A matrix containing proportions at each bin.

Wbin: A matrix containing surprisal values at each bin.

indfine: A vector of length 101 with equally spaced score values used for plotting.

Pmatfine: A matrix of probabilities over a fine mesh.

Wmatfine: A matrix of surprisal values over a fine mesh.

DWmatfine: A matrix of the values of the first derivative of surprisal curves over fine mesh.

D2Wmatfine: A matrix of the values of the second derivative of surprisal curves over fine mesh.

PStdErr: A matrix of the values of the standard error of probability curves at each bin.

WStdErr: A matrix of the values of the standard error of surprisal curves at each bin.

`dataList` A list that contains the objects needed to analyse the test or rating scale with the following fields:

U: A matrix of response data with N rows and n columns where N is the number of examinees or respondents and n is the number of items. Entries in the matrices are the indices of the options chosen. Column i of U is expected to contain only the integers `1,...,noption`.

optList: A list vector containing the numerical score values assigned to the options for this question.

key: If the data are from a test of the multiple choices type where the right answer is scored 1 and the wrong answers 0, this is a numeric vector of length n containing the indices the right answers. Otherwise, it is NULL.

WfdPar: An fdPar object for the defining the surprisal curves.

noption: A numeric vector of length n containing the numbers of options for each item.

nbin: The number of bins for binning the data.

scrrng: A vector of length 2 containing the limits of observed sum scores.

scrfine: A fine mesh of test score values for plotting.

scrvec: A vector of length N containing the examinee or respondent sum scores.

itemvec: A vector of length n containing the question or item sum scores.

percntrnk: A vector length N containing the sum score percentile ranks.

thetaQnt: A numeric vector of length 2*nbin + 1 containing the bin boundaries alternating with the bin centers. These are initially defined as `seq(0,100,len=2*nbin+1)`.

Wdim: The total dimension of the surprisal scores.

PcntMarkers: The marker percentages for plotting: 5, 25, 50, 75 and 95.

`Qvec` A vector of five marker percentile values. For plotting over information, this is replaced by Qvec_al returned as parList$Qvec_al.

`binctr` A vector of bin center values. If the plot is over arc length or information, `binctr` is modified before calling `Wbinsth.plot` by the command `binctr_al = pracma::interp1(indfine, alfine, binctr)`, and argument `binctr` is replaced by `binctr_al`.

`plotType` Type(s) of plot, default as "P" for probability, can also be "W" for surprisal, "DW" for sensitivity, and any combination of the three

`plotindex` A vector of indices of items to be plotted.

`plotrange` A vector of length 2 containing the plot boundaries of the score index interval.

`shaderange` a list of length 2 vector(s); set if users want to gray out specific score range(s)

`Wrng` A vector of length 2 specifing the plotting range for surprisal values.

`DWrng` A vector of length 2 specifing the plotting range for sensitivity values.

`titlestr` plot title

`autoplot` in Vignette, plot all items in a batch

`data_point` A logical value indicating whether to plot the data points.

`ci` A logical value indicating whether to plot the confidence limits.

`ttlsz` Title font size.

`axisttl` Axis title font size.

`axistxt` Axis text(tick label) font size.

`lgdlab` Legend label font size.

Value

A list vector is returned which is of the length of argument `plotindex`. Each member of the vector is a `gg` or `ggplot` object for the associated `plotindex` `value`. Each plot can be displayed using the `print` command. The plots of item power are produced as a side value even if no output object is specified in the call to the function.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

Sensitivity.plot, Power.plot, Entropy.plot, Wbinsmth,

Examples

```
#  Example 1.  Display the item surprisal curves for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
WfdList  <- Quantshort_parList$WfdList
Qvec     <- Quantshort_parList$Qvec
binctr   <- Quantshort_parList$binctr
titlestr <- NULL
#  plot the curves for the first question
scrfine <- seq(0,100,len=101)
oldpar <- par(no.readonly=TRUE)
ICC.plot(scrfine, WfdList, Quantshort_dataList, Qvec, binctr, plotindex=1:2)
par(oldpar)
#  Example 2.  Display the item proability and surprisal curves for the
# Symptom Distress Scale with 13 items and 473 respondents.
# Proceed as above changing "Quant" to "SDS"
```

---

keyshsort *Option information for the short form of the SweSAT Quantitative test.*

Description

The file has two rows, each with 24 integers in string format. The first row contains the indices of the right answers among the options for the 24 questions. The second row contains the number of options for each question.

Usage

```
keyshort
```

Details

The code above inputs the contents of the file using the scan function into a single character vector of length 48. The strings are converted to integers by the second command, and the integer vector is then reformatted into a 2 by 24 matrix. The subsequent analysis of these data is described in detail in the vignette SweSATQuanitativeShort.

`make.dataList` *Make a list object containing information required for analysis of choice data.*

Description

The returned named list object contains 11 named members, which are described in the value section below.

Usage

```
make.dataList(U, key, optList, grbg=rep(0,n), scrrng=NULL, titlestr=NULL,
              nbin=nbinDefault(N),NumBasis=NULL,  WfdPar=NULL,
              jitterwrd=TRUE, PcntMarkers=c( 5, 25, 50, 75, 95),
              quadwrd=FALSE, verbose=FALSE)
```

Arguments

`U`	A matrix with rows corresponding to examinees or respondents, and columns to questions or items.
`key`	If the data are a multiple choice test with only weights 0 and 1, a vector of length n containing the indices of the right answers. Otherwise, NULL.
`optList`	A list vector of length number of questions. A member contains a vector of score values assigned to each answer or option by the test designer.
`grbg`	A vector of length indicating which option for each item contains missing or illegal choice values. If 0, there is no such option.
`scrrng`	A vector of length two containing the initial and final values for the interval over which test scores are to be plotted. Default is minimum and maximum sum score.
`titlestr`	A string to be used as a title in plots and other displays.
`nbin`	The number of bins for containing proportions of examinees choosing options. The default is computed by a function that uses the number of examinees.
`NumBasis`	The number of spline basis functions used to represent surprisal curves. The default is computed by a function that uses the number of examinees.
`WfdPar`	A functional parameter object specifying a basis, a linear differential operator, a smoothing parameter, a boolian constant indicating estimation, and a penalty matrix. The default object defines an order 5 spline with operator `Lfd = 3` and smoothing parameter `1e4`. The number of basis functions depends on the size of N.
`jitterwrd`	A boolian constant: TRUE implies adding a small random value to each sum score value prior to computing percent rank values.
`PcntMarkers`	Used in plots of curves to display marker or reference percentage points for abscissa values in plots.

quadwrd	If TRUE a default order 5 basis is constructed. If FALSE the order and number of of basis functions are 3 so as to support the use of the second derivative in function thetafun.
verbose	If TRUE details of calculations are displayed.

Details

The score range defined scrrng should contain all of the sum score values, but can go beyond them if desired. For example, it may be that no examinee gets a zero sum score, but for reporting and display purposes using zero as the lower limit seems desirable.

The number of bins is chosen so that a minimum of at least about 25 initial percentage ranks fall within a bin. For larger samples, the number per bin is also larger, making the proportions of choice more accurate. The number bins can be set by the user, or by a simple algorithm used to adjust the number of bins to the number N or examinees.

The number of spline basis functions used to represent a surprisal curve should be small for small sample sizes, but can be larger when larger samples are involved. A minimum of 7 and maximum of 24 basis functions is assigned by default. The number of basis function can be set by the user or otherwise by a simple algorithm used to adjust the number of bins to the number N of examinees.

Adding a small value to discrete values before computing ranks is considered a useful way of avoiding any biasses that might arise from the way the data are stored. The small values used leave the rounded jittered values fixed, but break up ties for sum scores.

It can be helpful to see in a plot where special marker percentages 5, 25, 50, 75 and 95 percent of the interval [0,100] are located. The median abscissa value is at 50 per cent for initial percent rank values, for example, but may not be located at the center of the interval after iterations of the analysis cycle.

Value

A named list with named members as follows:

U: A matrix of response data with N rows and n columns where N is the number of examinees or respondents and n is the number of items. Entries in the matrices are the indices of the options chosen. Column i of U is expected to contain only the integers 1,...,noption.

optList: A list vector containing the numerical score values assigned to the options for this question.

key: If the data are from a test of the multiple choices type where the right answer is scored 1 and the wrong answers 0, this is a numeric vector of length n containing the indices the right answers. Otherwise, it is NULL.

WfdPar: An fdPar object for the defining the surprisal curves.

noption: A numeric vector of length n containing the numbers of options for each item.

nbin: The number of bins for binning the data.

scrrng: A vector of length 2 containing the limits of observed sum scores.

scrfine: A fine mesh of test score values for plotting.

scrvec: A vector of length N containing the examinee or respondent sum scores.

itemvec: A vector of length n containing the question or item sum scores.

percntrnk: A vector length N containing the sum score percentile ranks.

thetaQnt: A numeric vector of length 2*nbin + 1 containing the bin boundaries alternating with the bin centers. These are initially defined as `seq(0,100,len=2*nbin+1)`.

Wdim: The total dimension of the surprisal scores.

PcntMarkers: The marker percentages for plotting: 5, 25, 50, 75 and 95.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

Analyze

Examples

```
#  Example 1:  Input choice data and key for the short version of the
#  SweSAT quantitative multiple choice test with 24 items and 1000 examinees
#  input the choice data as 1000 strings of length 24
N <- dim(Quantshort_U)[1]
n <- dim(Quantshort_U)[2]

optList <- list()
grbg = matrix(0,n,1)
for (item in 1:n){
  noptioni <- length(unique(Quantshort_U[,item]))
  scorei <- rep(0,noptioni)
  scorei[Quantshort_key[item]] <- 1
  optList[[item]] <- scorei
  grbg[item] <- length(scorei)
}
optList <- list(itemLab=NULL, optLab=NULL, optScr=optList)

#  Set up the dataList object containing the objects necessary
#  for further display and analyses

Quantshort_dataList <- make.dataList(Quantshort_U, Quantshort_key, optList, grbg)

#  Example 2:  Input choice data and key for the Symptom Distress Scale
#  with 13 items and 473 examinees.
#  input the choice data as 473 strings of length 13
N <- dim(SDS_U)[1]
```

```
n <- dim(SDS_U)[2]
# --------- Define the option score values for each item ---------
optList <- list()
grbg <- matrix(0,n,1)
for (item in 1:n){
  scorei <- c(0:4,0)
  optList[[item]] <- scorei
  grbg[item] <- length(scorei)
}
optList <- list(itemLab=NULL, optLab=NULL, optScr=optList)
#  largest observed sum score is 37
scrrng <- c(0,37)
SDS_dataList <- make.dataList(SDS_U, SDS_key, optList, grbg, scrrng)
```

`mu.plot` *Plot expected test score as a function of score index*

Description

The expected score `mu(theta)` is a function of the score index theta. A diagonal dashed line is displayed to show the linear relationship to the score range interval.

Usage

```
mu.plot(mufine, scrrng, titlestr)
```

Arguments

`mufine`	A mesh of 101 equally spaced values of mu as a function of theta.
`scrrng`	A vector of length 2 containing the score range.
`titlestr`	A string containing the title of the data.

Value

A `gg` or `ggplot` object defining the plot of the expected test score `mu` as a function of the score index `theta`. This is displayed by the `print` command. The plot is automatically displayed as a side value even if no return object is specified in the calling statement.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

scoreDensity, testscore

Power.plot *Plot item power curves for selected items or questions.*

Description

Item surprisal power curves are the square root of the sum over options of the squared surprisal sensitivity curves.

Usage

```
Power.plot(scrfine, WfdList, Qvec, dataList, plotindex=1:n,
           plotrange=c(min(scrfine),max(scrfine)), height=0.5,
           value=0, ttlsz=NULL, axisttl=NULL, axistxt=NULL)
```

Arguments

scrfine	A vector of length nfine (usually 101) containing equally spaced points spanning the plotrange. Used for plotting.
WfdList	A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of WfdList is a named list containing information computed during the analysis. These named lists contain these objects: **Wfd:** A functional data object containing the M surprisal curves for a question. **M:** The number of options. **Pbin:** A matrix containing proportions at each bin. **Wbin:** A matrix containing surprisal values at each bin. **Pmatfine:** A matrix of probabilities over a fine mesh. **Wmatfine:** A matrix of surprisal values over a fine mesh. **DWmatfine:** A matrix of the values of the first derivative of surprisal curves over fine mesh. **D2Wmatfine:** A matrix of the values of the second derivative of surprisal curves over fine mesh.
Qvec	The five marker percentile values.
dataList	A list vector containing objects essential to an analysis.
plotindex	A set of integers specifying the numbers of the items or questions to be displayed.
plotrange	A vector of length 2 containing the plot boundaries within or over the score index interval c(0,100).
height	A positive real number defining the upper limit on the ordinate for the plots.
value	Number required by ggplot2. Defaults to 0.

ttlsz	Title font size.
axisttl	Axis title font size.
axistxt	Axis text(tick label) font size.

Details

A surprisal power curve for each question indexed in the index argument. A request for a keystroke is made for each question. The answer to question strongly defines the optimal position of an estimated score index value where the curve is high value. Values of power curves typically range over [0,0.5].

Value

The plots of the power curves specified in plotindex are produced as a side effect. If saveplot is TRUE, the plots of item power curves specified in plotindex are bundled into a single postscript or .pdf file and the file name is defined by paste(dataList$titlestr,i,'-power.pdf',sep="").
The file is then outputas a returned value.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

Sensitivity.plot, Entropy.plot, Hfuns.plot, ICC.plot

Examples

```
#  Example 1.  Display the item power curves for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
#  plot the power curve for the first item
WfdList <- Quantshort_parList$WfdList
Qvec    <- Quantshort_parList$Qvec
scrfine <- seq(0,100,len=101)
oldpar <- par(no.readonly=TRUE)
Power.plot(scrfine, WfdList, Qvec, Quantshort_dataList, plotindex=1)
par(oldpar)
#  Example 2.  Display the item power curves for the
#  Symptom Distress Scale with 13 items and 473 respondents
#  Proceed as above changing "Quant" to "SDS".
```

`Quantshort_dataList` *List of objects essential for an analysis of the abbreviated SweSAT Quantitative multiple choice test.*

Description

The data are for 1000 randomly selected examinees taking 24 math analysis multiple choice questions.

Usage

```
Quantshort_dataList
```

Format

A named list.

Details

A named list with 15 members:

U: A matrix of response data with N rows and n columns where N is the number of examinees or respondents and n is the number of items. Entries in the matrices are the indices of the options chosen. Column i of U is expected to contain only the integers `1,...,noption`.

optList: A list vector containing the numerical score values assigned to the options for this question.

key: If the data are from a test of the multiple choices type where the right answer is scored 1 and the wrong answers 0, this is a numeric vector of length n containing the indices the right answers. Otherwise, it is NULL.

grbg: A vector of length indicating which option for each item contains missing or illegal choice values. If 0, there is no such option.

WfdPar: An fdPar object for the defining the surprisal curves.

noption: A numeric vector of length n containing the numbers of options for each item.

nbin: The number of bins for binning the data.

scrrng: A vector of length 2 containing the limits of observed sum scores.

scrfine: A fine mesh of test score values for plotting.

scrvec: A vector of length N containing the examinee or respondent sum scores.

itemvec: A vector of length n containing the question or item sum scores.

percntrnk: A vector length N containing the sum score percentile ranks.

thetaQnt: A numeric vector of length 2*nbin + 1 containing the bin boundaries alternating with the bin centers. These are initially defined as `seq(0,100,len=2*nbin+1)`.

Wdim: The total dimension of the surprisal scores.

PcntMarkers: The marker percentages for plotting: 5, 25, 50, 75 and 95.

titlestr: A string containing a title for the analysis.

`Quantshort_infoList` *Arclength or information arameter list for 24 items from the quantitative SweSAT subtest.*

Description

The data are for 1000 examinees randomly selected from those who took the 2013 quantitative subtest of the SweSAT university entrance exam. The questions are only the 24 math analysis questions, and each question has four options. The analysis results are after 10 cycles of alternating between estimating surprisal curves and estimating percentile score index values. The objects in list object `Quantshort_infoList` are required for plotting results over the arc length or information domain rather the score index domain. This domain is preferred because such plots are invariant with respect to changes in the score index domain. It also has a metric structure so that differences are comparable no matter where they fall within the information domain.

Usage

```
Quantshort_infoList
```

Format

A named list containing eight objects.

Value

The object `Quantshort_parList` is a named list with these members:

arclength: The total length of the information domain measured in M-bits, where `M` is the number of options for a question.

Wfd.theta: The log derivative functional data object defining a strictly increasing set of arc length values corresponding to set of score index values.

arclengthvec: A mesh of equally-spaced values of indefinite integrals of sum of norms of surprisal derivatives.

theta_al The N arc length values corresponding to the N estimated score index values assigned to N examinees.

Qvec_al: The arc length positions corresponding to the marker percentages 5, 25, 50, 75 and 95.

Wfd.info: The log derivative functional data object defining a strictly increasing set of score index valuescorresponding to a set of arc length values.

thetavec: A vector of score index values resulting from using function `monfd` with equally spaced arc length values and `Wfd.info`.

Wdim: The dimension of the over space containing the surprisal pcurves.

Quantshort_key *Option information for the short form of the SweSAT Quantitative test.*

Description

A vector that contains the indices of the right answers among the options for the 24 questions

Usage

```
Quantshort_key
```

Quantshort_parList *Parameter list for 24 items from the quantitative SweSAT subtest.*

Description

The data are for 1000 examinees randomly selected from those who took the 2013 quantitative subtest of the SweSAT university entrance exam. The questions are only the 24 math analysis questions, and each question has four options. The analysis results are after 10 cycles of alternating between estimating surprisal curves and estimating percentile score index values.

Usage

```
Quantshort_parList
```

Format

A named list.

Value

The object `Quantshort_parList` is a named list with these members:

theta: A vector of length N of estimated values of the percentile rank score index.

thetaQnt: A vector of length 2*nbin + 1 containing bin boundaries alternating with bin centres.

WfdList: A list vector of length equal to the number of questions. Each member contains the following eight results for the surprisal curves associated with a question:

- **Wfd:** A functional data object representing the M surprisal curves defined over the percentile rank range, where M is the number of options for that question.
- **M:** The number of option choices for the question.
- **type:** The type of question: 1 for multiple choice, 0 for rating scale.
- **Pbin:** A nbin by M matrix of proportions of examinees choosing each option for each bin,.
- **Wbin:** A nbin by M matrix of surprisal values of examinees choosing each option for each bin.

Pmatfine: A nfine by M matrix of proportions over a fine mesh of proportions of examinees choosing each option.

Wmatfine: A nfine by M matrix of surprisal values over a fine mesh of values of examinees choosing each option.

DWmatfine: A nfine by M matrix of the first derivative of surprisal values over a fine mesh of values of examinees choosing each option. These are referred to as the sensitivity values.

logdensfd: A functional data object representing the logarithm of the density of the percentile rank score index values.

C: The norming constant: the density function is `exp(logdensfd)/C`.

densfine: A fine mesh of probability density values of the percentile rank score index.

denscdf: A fine mesh of cumulative probability distribution values used for interpolating values.

Qvec: The score index values associated with the five marker percentages 5, 25, 50, 75 and 95.

binctr: A vector of length nbin containing the centres of the bins.

bdry: A vector of length nbin+1 containing the boundaries of the bins.

freq: An nbin by M matrix of frequencies with which options are chosen.

Wmax: A maximum surprisal value used for plotting purposes.

Hval: The value of the fitting criterion `H` for a single examinee or respondent.

DHval: The value of the first derivative of the fitting criterion `H` for a single examinee or respondent.

D2Hval: The value of the second derivative of the fitting criterion `H` for a single examinee or respondent.

active: A logical vector of length N indicating which estimates of theta are converged (FALSE) or not converged (TRUE).

arclength: The length in bits of the test information curve.

alfine: A mesh of 101 equally spaced positions along the test information curve.

Qvec_al: The positions of the five marker percentages on the test information curve.

theta_al: A vector of length N containing the positions of each examinee or respondent on the test information curve.

`Quantshort_U` *Test data for 24 math calculation questions from the SweSAT Quantitative data.*

Description

These data are for a randomly selected subset of 1000 examinees.

Usage

```
Quantshort_U
```

Format

A matrix object with 1000 rows and 24 columns. The integers indicate which answer was chosen for each question by the examinee associated with the row.

scoreDensity *Compute and plot a score density histogram and and curve.*

Description

The tasks of function `theta.density()` and plotting the density are combined. The score density is plotted both as a histogram and as a smooth curve. All the score types may be plotted: sum scores, expected test scores, percentile score index values, and locations on the test information or scale curve. The plot is output as a ggplot2 plot object, which is actually plotted using the `print` command.

Usage

```
scoreDensity(scrvec, scrrng=c(0,100), ndensbasis=15, ttlstr=NULL, pltmax=0)
```

Arguments

`scrvec`	A vector of strictly increasing bin boundary values, with the first at the lowest plotting value and the last at the upper boundary. The number of bins in the histogram is one less than the number of bndry values.
`scrrng`	A vector of length 2 containing lower and upper boundaries on scores, which defaults to c(0,100).
`ndensbasis`	The number of spline basis functions to be used to represent the smooth density curve.
`ttlstr`	A string object used as a title for the plot. Defaults to none.
`pltmax`	An upper limit on the vertical axis for plotting. Defaults to the maximum curve value.

Value

A ggplot2 plot object `dens.plot` that can be displayed using command `print(dens.plot)`.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

`thetafun`, `theta2arclen`, `testscore`, `theta.distn`

Examples

```
#  Example 1.  Display probability density curves for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
WfdList <- Quantshort_parList$WfdList
theta   <- Quantshort_parList$theta
Qvec    <- Quantshort_parList$Qvec
#  plot the density for the score indices within interval c(0,100)
theta_int <- theta[0 < theta & theta < 100]
oldpar <- par(no.readonly=TRUE)
scoreDensity(theta_int)
par(oldpar)
```

`scorePerformance` *Calculate mean squared error and bias for a set of score index values from simulated data.*

Description

After the simulated data matrices have been analyzed, prepare the objects necessary for the performance plots produced by functions `RMSEbias1.plot` and `RMSEbias2.plot`.

Usage

```
scorePerformance(dataList, simList)
```

Arguments

`dataList` A list that contains the objects needed to analyse the test or rating scale with the following fields:

U: A matrix of response data with N rows and n columns where N is the number of examinees or respondents and n is the number of items. Entries in the matrices are the indices of the options chosen. Column i of U is expected to contain only the integers `1,...,noption`.

optList: A list vector containing the numerical score values assigned to the options for this question.

key: If the data are from a test of the multiple choices type where the right answer is scored 1 and the wrong answers 0, this is a numeric vector of length n containing the indices the right answers. Otherwise, it is NULL.

WfdPar: An fdPar object for the defining the surprisal curves.

noption: A numeric vector of length n containing the numbers of options for each item.

nbin: The number of bins for binning the data.

scrrng: A vector of length 2 containing the limits of observed sum scores.

scrfine: A fine mesh of test score values for plotting.

scrvec: A vector of length N containing the examinee or respondent sum scores.

itemvec: A vector of length n containing the question or item sum scores.

percntrnk: A vector length N containing the sum score percentile ranks.

thetaQnt: A numeric vector of length 2*nbin + 1 containing the bin boundaries alternating with the bin centers. These are initially defined as `seq(0,100,len=2*nbin+1)`.

Wdim: The total dimension of the surprisal scores.

PcntMarkers: The marker percentages for plotting: 5, 25, 50, 75 and 95.

`simList` A named list containing these objects:

sumscr: A matrix with row dimension `ntheta`, the number of population score index values and column dimension `nsample`, the number of simulated samples.

theta: An `ntheta` by `nsample` of estimated score index values.

mu: An `ntheta` by `nsample` of estimated expected score values.

al: An `ntheta` by `nsample` of estimated test information curve values.

thepop: A vector of population score index values.

mupop: A vector of expected scores computed from the population score index values.

alpop: A vector of test information values computed from the population score index values.

n: The number of questions.

Qvec: The five marker percentile values.

Value

A named list containing these objects:

sumscr: A matrix with row dimension `ntheta`, the number of population score index values and column dimension `nsample`, the number of simulated samples.

theta: An `ntheta` by `nsample` matrix of estimated score index values.

mu: An `ntheta` by `nsample` matrix of estimated expected score values.

al: An `ntheta` by `nsample` matrix of estimated test information curve values.

thepop: A vector of population score index values.

mupop: A vector of expected scores computed from the population score index values.

alpop: A vector of test information values computed from the population score index values.

n: The number of questions.

Qvec: The five marker percentile values.

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

`dataSimulation`

SDS *The Symptom Distress Scale data.*

Description

The SDS is a rating scale with 13 questions and scores 0, 1, 2, 3 and 4.

Usage

```
data("SDS")
```

Format

A text file with 473 rows, each containing a string of 13 characters.

SDS_dataList *List of objects essential for an analysis of the Symptom Distress Scale.*

Description

The data are for 473 respondents to the Symptom Distress Scale. Each question has rating scores 0, 1, 2, 3, and 4.

Usage

```
data("SDS_dataList")
```

Format

A named list.

Details

A named list with 15 members. See below for a description of each member.

U: An N by n matrix of integers indicating which option was chosen by one of N examinees for a question. The object should be in the matrix class and the contents should be integers. For any question the integers should be within the range 1 to the number of options.

optList: A list vector object of length equal to the number of questions. Each member of the list is a numeric vector of score values assigned to each option by the test designer. For multiple choice questions, these scores are usually 0 for wrong answers and 1 for the right answer. For rating scales they are often a sequence of signed integers indicating intensity of some experience.

grbg: A vector of length indicating which option for each item contains missing or illegal choice values. If 0, there is no such option.

WfdPar: A functional parameter object that is required to represent the surprisal curves. The number of basis functions should be small for smaller sizes and larger for large samples. For example, 7 for N <= 500, 11 for N in [500,2000], 24 for N >= 2000.

logdensfd: A functional data object used to represent the log density of the percentile score index values. The number of basis functions is often the same as used for the WfdPar object.

noption: An integer vector of length equal to the number of questions indicating the number of options for each question.

nbin: The number of bins used to contain the score index values. The number should ensure that each bin can hold at least 25 score values. For medium sized samples, 50 per bin is fine, and for large samples as many as are required to define the detail in the surprisal curves.

key: An integer vector of length equal to the number of questions and containing the right answer indices. This is not used for rating scales and can be NULL in that case.

scrrng: A vector of length 2 defining a range over which the sum and expected scores will be estimated.

scrvec: A vector of length N containing the sum scores for the examinees or respondents.

percntrnk: The percentile ranks for each of the examinees or respondents. This is the percentage of examinees with sum scores below that for an examinee. Ties are often broken by adding a small random quantity to each sum score before computing ranks.

scrfine: A fine mesh of score values used for plotting purposes, and running from scrrng[1] to scrrng[2]. Common number for mesh values are 51 or 101.

scrtype: If the question is of multiple choice, the value is 1, and if not, 0.

indrng: Defines the range of the score index. Usually this is c(0,100).

indfine: A fine mesh of score index values used for plotting, and usually ranging from 0 to 100. The number of mesh values is commonly 51 or 101.

`SDS_infoList` *Arclength or information parameter list for 13 items in the Symptom Distress Scale.*

Description

The data are for 473 respondents to the Symptom Distress Scale. Each question has options with rating scores 0, 1, 2, 3, and 4. The analysis results are 10 cycles of alternating between estimating surprisal curves and estimating percentile score index values.

The objects in list object `Quantshort_infoList` are required for plotting results over the arc length or information domain rather the score index domain. This domain is preferred because such plots are invariant with respect to changes in the score index domain. It also has a metric structure so that differences are comparable no matter where they fall within the information domain.

Usage

```
SDS_infoList
```

Format

A named list containing eight objects.

Value

The object `Quantshort_parList` is a named list with these members:

arclength: The total length of the information domain measured in M-bits, where `M` is the number of options for a question.

Wfd.theta: The log derivative functional data object defining a strictly increasing set of arc length values corresponding to set of score index values.

arclengthvec: A mesh of equally-spaced values of indefinite integrals of sum of norms of surprisal derivatives.

theta_al The N arc length values corresponding to the N estimated score index values assigned to N examinees.

Qvec_al: The arc length positions corresponding to the marker percentages 5, 25, 50, 75 and 95.

Wfd.info: The log derivative functional data object defining a strictly increasing set of score index valuescorresponding to a set of arc length values.

thetavec: A vector of score index values resulting from using function `monfd` with equally spaced arc length values and `Wfd.info`.

Wdim: The dimension of the over space containing the surprisal curves.

`SDS_key` *Key for Symptom Distress Scale.*

Description

NULL for scale data.

Usage

```
SDS_key
```

SDS_parList *Parameter list for the Symptom Distress Scale.*

Description

The data are for 473 respondents to the Symptom Distress Scale. Each question has options with rating scores 0, 1, 2, 3, and 4. The analysis results are 10 cycles of alternating between estimating surprisal curves and estimating percentile score index values.

Usage

```
SDS_parList
```

Format

A named list.

Value

The object `SDS_dataList` is a list is a named list with these members:

theta: A vector of length N of estimated values of the percentile rank score index.

thetaQnt: A vector of length 2*nbin + 1 containing bin boundaries alternating with bin centres.

WfdList: A list vector of length equal to the number of questions. Each member contains the following eight results for the surprisal curves associated with a question:

- Wfd:A functional data object representing the M surprisal curves defined over the percentile rank range, where M is the number of options for that question.
- M:The number of option choices for the question.
- type:The type of question: 1 for multiple choice, 0 for rating scale.
- Pbin:A nbin by M matrix of proportions of examinees choosing each option for each bin,.
- Wbin:A nbin by M matrix of surprisal values of examinees choosing each option for each bin.
- Pmatfine:A nfine by M matrix of proportions over a fine mesh of proportions of examinees choosing each option.
- Wmatfine:A nfine by M matrix of surprisal values over a fine mesh of values of examinees choosing each option.
- DWmatfine:A nfine by M matrix of the first derivative of surprisal values over a fine mesh of values of examinees choosing each option. These are referred to as the sensitivity values.

logdensfd: A functional data object representing the logarithm of the density of the percentile rank score index values.

C: The norming constant: the density function is `exp(logdensfd)/C`.

densfine: A fine mesh of probability density values of the percentile rank score index.

denscdf: A fine mesh of cumulative probability distribution values used for interpolating values.

Qvec: The score index values associated with the five marker percentages 5, 25, 50, 75 and 95.

binctr: A vector of length nbin containing the centres of the bins.

bdry: A vector of length nbin+1 containing the boundaries of the bins.

freq: An nbin by M matrix of frequencies with which options are chosen.

Wmax: A maximum surprisal value used for plotting purposes.

Hval: The value of the fitting criterion H for a single examinee or respondent.

DHval: The value of the first derivative of the fitting criterion H for a single examinee or respondent.

D2Hval: The value of the second derivative of the fitting criterion H for a single examinee or respondent.

active: A logical vector of length N indicating which estimates of theta are converged (FALSE) or not converged (TRUE).

arclength: The length in bits of the test information curve.

alfine: A mesh of 101 equally spaced positions along the test information curve.

Qvec_al: The positions of the five marker percentages on the test information curve.

theta_al: A vector of length N containing the positions of each examinee or respondent on the test information curve.

SDS_U *Test data for Symptom Distress Scale.*

Description

The integers indicate which answer was chosen for each question by the examinee associated with the row.

Usage

```
SDS_U
```

Format

A vector object with 473 rows and 13 columns.

`Sensitivity.plot` *Plots all the sensitivity curves for selected items or questions.*

Description

A sensitivity curve for an option is the first derivative of the corresponding surprisal curve. Its values can be positive or negative, and the size of the departure from zero at any point on the curve is the amount information contributed by that curve to locating the value of an examinee or respondent on the score index continuum.

Usage

```
Sensitivity.plot(scrfine, WfdList, Qvec, dataList, plotindex=1:n,
                 plotrange=c(min(scrfine),max(scrfine)),
                 key=NULL, titlestr=NULL, saveplot=FALSE, width=c(-0.2,0.2),
                 ttlsz=NULL, axisttl=NULL, axistxt=NULL, lgdlab=NULL)
```

Arguments

`scrfine` A vector of length nfine (usually 101) containing equally spaced points spanning the plotrange. Used for plotting.

`WfdList` A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of `WfdList` is a named list containing information computed during the analysis. These named lists contain these objects:

Wfd: A functional data object containing the `M` surprisal curves for a question.

M: The number of options.

Pbin: A matrix containing proportions at each bin.

Wbin: A matrix containing surprisal values at each bin.

Pmatfine: A matrix of probabilities over a fine mesh.

Wmatfine: A matrix of surprisal values over a fine mesh.

DWmatfine: A matrix of the values of the first derivative of surprisal curves over fine mesh.

D2Wmatfine: A matrix of the values of the second derivative of surprisal curves over fine mesh.

`Qvec` The values of the five marker percentiles.

`dataList` A list that contains the objects needed to analyse the test or rating scale with the following fields:

U: A matrix of response data with N rows and n columns where N is the number of examinees or respondents and n is the number of items. Entries in the matrices are the indices of the options chosen. Column i of U is expected to contain only the integers 1,...,noption.

optList: A list vector containing the numerical score values assigned to the options for this question.

key: If the data are from a test of the multiple choices type where the right answer is scored 1 and the wrong answers 0, this is a numeric vector of length n containing the indices the right answers. Otherwise, it is NULL.

WfdPar: An fdPar object for the defining the surprisal curves.

noption: A numeric vector of length n containing the numbers of options for each item.

nbin: The number of bins for binning the data.

scrrng: A vector of length 2 containing the limits of observed sum scores.

scrfine: A fine mesh of test score values for plotting.

scrvec: A vector of length N containing the examinee or respondent sum scores.

itemvec: A vector of length n containing the question or item sum scores.

percntrnk: A vector length N containing the sum score percentile ranks.

thetaQnt: A numeric vector of length 2*nbin + 1 containing the bin boundaries alternating with the bin centers. These are initially defined as seq(0,100,len=2*nbin+1).

Wdim: The total dimension of the surprisal scores.

PcntMarkers: The marker percentages for plotting: 5, 25, 50, 75 and 95.

`plotindex` A set of integers specifying the numbers of the items or questions to be displayed.

`plotrange` A vector of length 2 containing the plot boundaries within or over the score index interval c(0,100).

`key` A integer vector of indices of right answers. If the data are rating scales, this can be NULL.

`titlestr` A title string for plots.

`saveplot` A logical value indicating whether the plot should be saved to a pdf file.

`width` A vector of length 2 defining the lower and upper limits on the ordinate for the plots.

`ttlsz` Title font size.

`axisttl` Axis title font size.

`axistxt` Axis text(tick label) font size.

`lgdlab` Legend label font size.

Details

Sensitivity curves for each question indexed in the `index` argument. A request for a keystroke is made for each question.

Value

A list vector is returned which is of the length of argument `plotindex`. Each member of the vector is a `gg` or `ggplot` object for the associated `plotindex` `value`. Each plot can be displayed using the `print` command. The plots of item power are produced as a side value even if no output object is specified in the call to the function.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

Power.plot, Entropy.plot, ICC.plot

Examples

```
#  Example 1.  Display the option sensitivity curves for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
WfdList <- Quantshort_parList$WfdList
Qvec    <- Quantshort_parList$Qvec
scrfine <- seq(0,100,len=101)
oldpar <- par(no.readonly=TRUE)
Sensitivity.plot(scrfine, WfdList, Qvec, Quantshort_dataList,
                 plotindex=1)
par(oldpar)
#  Example 2.  Display the option sensitivity curves for the
#  Symptom Distress Scale with 13 items and 473 respondents
#  Proceed as above changing "Quant" to "SDS".
```

`testscore` *Compute the expected test score by substituting probability of choices for indicator variable 0-1 values. Binary items assumed coded as two choice items.*

Description

Compute the expected test score by substituting probability of choices for indicator variable 0-1 values. Binary items assumed coded as two choice items.

Usage

```
testscore(theta, WfdList, optList)
```

Arguments

theta	Initial values for score indices in the interval [0,100]. A vector of size N.
WfdList	A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of WfdList is a named list containing information computed during the analysis. These named list in each member contains these objects: **Wfd:** A functional data object containing the M surprisal curves for a question. **M:** The number of options. **Pbin:** A matrix containing proportions at each bin. **Wbin:** A matrix containing surprisal values at each bin. **Pmatfine:** A matrix of probabilities over a fine mesh. **Wmatfine:** A matrix of surprisal values over a fine mesh. **DWmatfine:** A matrix of the values of the first derivative of surprisal curves over fine mesh. **D2Wmatfine:** A matrix of the values of the second derivative of surprisal curves over fine mesh.
optList	A numbered list of length n. Each member contains the weights assigned to each option for that item or question.

Value

A vector of test score values.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

scoreDensity

Examples

```
#  Example 1.  Compute expected sum score values for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
WfdList <- Quantshort_parList$WfdList
theta   <- Quantshort_parList$theta
mu      <- testscore(theta, WfdList, Quantshort_dataList$optList)
par(c(1,1))
```

```
hist(mu,11)
#  Example 2.  Compute expected sum score values for the
#  Symptom Distress Scale with 13 items and 473 respondents
#  Proceed as above changing "Quant" to "SDS".
```

TG_density.fd *Compute a Probability Density Function*

Description

Like the regular S-PLUS function density, this function computes a probability density function for a sample of values of a random variable. However, in this case the density function is defined by a functional parameter object logdensfdPar along with a normalizing constant C.

The density function $p(thetadens)$ has the form p(thetadens) = C exp[W(thetadens)] where function $W(thetadens)$ is defined by the functional data object logdensfdPar.

Usage

```
## S3 method for class 'fd'
TG_density(thetadens, logdensfdPar, conv=0.0001, iterlim=20,
           active=1:nbasis, dbglev=0)
```

Arguments

thetadens	a set observations, which may be one of two forms: 1. a vector of observations $thetadens_i$ 2. a two-column matrix, with the observations $thetadens_i$ in the first column, and frequencies f_i in the second. The first option corresponds to all $f_i = 1$.
logdensfdPar	a functional parameter object specifying the initial value, basis object, roughness penalty and smoothing parameter defining function $W(t).$
conv	a positive constant defining the convergence criterion.
iterlim	the maximum number of iterations allowed.
active	a logical vector of length equal to the number of coefficients defining Wfdobj. If an entry is TRUE, the corresponding coefficient is estimated, and if FALSE, it is held at the value defining the argument Wfdobj. Normally the first coefficient is set to 0 and not estimated, since it is assumed that $W(0) = 0$.
dbglev	either 0, 1, or 2. This controls the amount information printed out on each iteration, with 0 implying no output, 1 intermediate output level, and 2 full output. If levels 1 and 2 are used, it is helpful to turn off the output buffering option in S-PLUS.

Details

The goal of the function is provide a smooth density function estimate that approaches some target density by an amount that is controlled by the linear differential operator `Lfdobj` and the penalty parameter. For example, if the second derivative of $W(t)$ is penalized heavily, this will force the function to approach a straight line, which in turn will force the density function itself to be nearly normal or Gaussian. Similarly, to each textbook density function there corresponds a $W(t)$, and to each of these in turn their corresponds a linear differential operator that will, when apply to $W(t)$, produce zero as a result. To plot the density function or to evaluate it, evaluate `Wfdobj`, exponentiate the resulting vector, and then divide by the normalizing constant `C`.

Value

a named list of length 4 containing:

`Wfdobj`	a functional data object defining function $W(thetadens)$ that that optimizes the fit to the data of the monotone function that it defines.
`C`	the normalizing constant.
`Flist`	a named list containing three results for the final converged solution: (1) **f**: the optimal function value being minimized, (2) **grad**: the gradient vector at the optimal solution, and (3) **norm**: the norm of the gradient vector at the optimal solution.
`iternum`	the number of iterations.
`iterhist`	a `iternum+1` by 5 matrix containing the iteration history.

See Also

`intensity.fd`

Examples

```
#  set up range for density
rangeval <- c(-3,3)
#  set up some standard normal data
thetadens <- rnorm(50)
#  make sure values within the range
thetadens[thetadens < -3] <- -2.99
thetadens[thetadens >  3] <-  2.99
#  set up basis for W(thetadens)
basisobj <- create.bspline.basis(rangeval, 11)
#  set up initial value for Wfdobj
Wfd0 <- fd(matrix(0,11,1), basisobj)
logdensfdPar <- fdPar(Wfd0)
#  estimate density
denslist <- TG_density.fd(thetadens, logdensfdPar)
#  plot density
thetadensval <- seq(-3,3,.2)
wval <- eval.fd(thetadensval, denslist$Wfdobj)
pval <- exp(wval)/denslist$C
oldpar <- par(no.readonly=TRUE)
```

```
on.exit(oldpar)
plot(thetadensval, pval, type="l", ylim=c(0,0.4))
points(thetadens,rep(0,50))
```

`theta.distn` *Compute score density*

Description

Computes the cumulated density for distribution function, the probability density function, and the log probability density function as fd objects by spline smoothing of the score values `thetadens` using the basis object `logdensbasis`. The norming constant `C` is also output.

The score values may score index values `theta`, expected test score values `mu`, or arc length locations on the test information or scale curve. The argument functional data object `logdensfd` should have a range that is appropriate for the score values being represented: For score indices, [0,100], for expected test scores, the range of observed or expected scores; and for test information curve locations in the interval [0,`arclength`].

Usage

```
theta.distn(thetadens, logdensbasis, pvec=c(0.05, 0.25, 0.50, 0.75, 0.95),
            nfine = 101)
```

Arguments

`thetadens`	A vector of score index, test score, or arc length values. In the score index case, these are usually only the values in the interior of the interval [0,100].
`logdensbasis`	A functional basis object for representing the log density function. The argument may also be a functional parameter object (`fdPar`) or a functional basis object (`Wbasis`).
`pvec`	A vector length NL containing the marker percentages.
`nfine`	The number of values in a fine grid, default as 101.

Value

A named list containing:

cdf_fd: An fd object for the cumulative probability function.

pdf_fd: An fd object for the probability density function. values over the fine mesh.

`logdensfd:`	A functional data object (`fd`) representing the log of the probability function for input `theta`.
`C:`	The normalization constant for computing the probability density function with the command `densityfd = exp(logdensfd)/C`.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

thetafun, theta2arclen, testscore, scoreDensity

Examples

```
#  Example 1.  Display the item power curves for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
#  Assemble information for estimating theta density
indfine   <- seq(0,100,len=101)
WfdList   <- Quantshort_parList$WfdList
theta     <- Quantshort_parList$theta
N         <- length(theta)
#  Define the density for only interior theta values
inside    <- theta > 0 & theta < 100
logdensbasis    <- Quantshort_parList$logdensfd$basis
theta.distnList <- theta.distn(theta[inside], logdensbasis)
theta.densfine  <- eval.fd(indfine, theta.distnList$pdf_fd)
#  Add number of values at each boundary
thetazero       <- length(theta[theta == 0])
thetacent       <- length(theta[theta == 100])
#  Plot both the interior density and the boundary proportions
oldpar <- par(no.readonly=TRUE)
on.exit(oldpar)
plot(indfine, theta.densfine, type="l", lwd=2, ylim=c(0,0.025),
     xlab="Percentile Index", ylab="Density")
lines(c(  0,100), c(1/100,1/100), lty=2)
lines(c(  0,  0), c(0,thetazero/N), lty=1, lwd=4)
lines(c(100,100), c(0,thetacent/N), lty=1, lwd=4)
```

---

theta2arclen *Compute results using arc length or information as the abscissa.*

Description

The one-dimensional psychometric model defines a space curve within the vector space defined by the total collection of option surprisal curves. This curve is a valuable resource since positions along the curve are defined in bits and positions on the curve are subject to the same strict properties that apply to physical measurements.

Function `theta2arclen` is required to convert objects defined over the score index continuum `c(0,100)` to the same objects over the arc length continuum `c(0,arclength)`, and also vice versa. Since the arc length or information continuum is along a space curve that is invariant under strictly monotone transformations of the score index `theta`, and is also a metric, it is an ideal choice for the abscissa in all plots.

Usage

```
theta2arclen(theta, Qvec, WfdList, binctr, itemindex=1:n, plotrng=c(0,100), shortwrd)
```

Arguments

`theta`	A vector of score index, test score, or arc length values, one for each examinee or respondent.
`Qvec`	A vector of locations of the five marker percentages.
`WfdList`	A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of `WfdList` is a named list containing information computed during the analysis. These named lists contain these objects: **Wfd:** A functional data object containing the `M` surprisal curves for a question. **M:** The number of options. **Pbin:** A matrix containing proportions at each bin. **Wbin:** A matrix containing surprisal values at each bin. **Pmatfine:** A matrix of probabilities over a fine mesh. **Wmatfine:** A matrix of surprisal values over a fine mesh. **DWmatfine:** A matrix of the values of the first derivative of surprisal curves over fine mesh. **D2Wmatfine:** A matrix of the values of the second derivative of surprisal curves over fine mesh.
`binctr`	A vector of locations of the bin centers.
`itemindex`	A vector containing the indices of the items to be used.
`plotrng`	A vector of length 2 containing the starting score index and end score index values of the range to be plotted.
`shortwrd`	If TRUE only vectors arclength and arclengtvec are returned in order to speed up the computation within cycles in function Analyze where only these objects are required. The default is FALSE.

Value

A named list object containing these results of the analysis:

arclength The length of the test information or scale curve.

arclengthvec Positions on the test information or scale curve corresponding to a fine mesh of score index values (typically 101 values between 0 and 100).

arclengthfd Functional data object representing the relation between the score index abscissa and the arclength or information ordinate.

theta_al A vector of positions on the test information or scale curve corresponding to the input score index values in argument `theta`.

Qvec_al Values in arc length of the five marker percentages.

binctr_al Values in arc length of the bin centers.

Wfd.info A functional data object representing the relation between the arclength or information abscissa and the score index ordinate.

Wdim.index The dimension of the overspace, which equal to sum of the number of options in the items specified in `itemindex`.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

`Wpca.plot`

Examples

```
#  Example 1.  Display the arc length curve for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
indfine       <- seq(0,100,len=101)
WfdList       <- Quantshort_parList$WfdList
theta         <- Quantshort_parList$theta
Qvec          <- Quantshort_parList$Qvec
binctr        <- Quantshort_parList$binctr
arclenList    <- theta2arclen(theta, Qvec, WfdList, binctr)
names(arclenList)
oldpar <- par(no.readonly=TRUE)
ArcLength.plot(arclenList$arclength, arclenList$arclengthvec)
#  Example 2.  Display the arc length curve for the
```

```
#  Symptom Distress Scale with 13 items and 473 respondents
#  Proceed as above changing "Quant" to "SDS".
par(oldpar)
```

---

thetafun *Compute optimal scores*

Description

The percentile score index values are estimated for each person. The estimates minimize the negative log likelihoods, which are a type of surprisal. The main optimization method is a safe-guarded Newton-Raphson method.

For any iteration the method uses only those scores that are within the interior of the interval [0,100] or at a boundary with a first derivative that would take a step into the interior, and have second derivative values exceeding the value of argument `crit`. Consequently the number of values being optimized decrease on each iteration, and iterations cease when either all values meet the convergence criterion or are optimized on a boundary, or when the number of iterations reaches `itermax`. At that point, if there are any interior scores still associated with either non-positive second derivatives or values that exceed `crit`, the minimizing value along a fine mesh is used.

If `itdisp` is positive, the number of values to be estimated are printed for each iteration.

Usage

```
thetafun(theta, WfdList, U, itermax = 20, crit = 0.001,
         itdisp = FALSE)
```

Arguments

`theta`	A vector of size `N` containing initial values for score indices in the interval [0,100].
`WfdList`	A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of `WfdList` is a named list containing information computed during the analysis. The named list in each member contains these objects: • Wfd: A functional data object containing the `M` surprisal curves for a question. • M: The number of options. • Pbin: A matrix containing proportions at each bin. • Wbin: A matrix containing surprisal values at each bin. • Pmatfine: A matrix of probabilities over a fine mesh. • Wmatfine: A matrix of surprisal values over a fine mesh. • DWmatfine: A matrix of the values of the first derivative of surprisal curves over fine mesh. • D2Wmatfine: A matrix of the values of the second derivative of surprisal curves over fine mesh.

U	A matrix number of rows equal to the number of examinees or respondents, and number of columns equal to number of items. The values in the matrix are indices of choices made by each respondent to each question.
itermax	Maximum number of iterations for computing the optimal theta values. Default is 20.
crit	Criterion for convergence of optimization. Default is 1e-8.
itdisp	If TRUE, results are displayed for each iteration.

Value

A named list with these members:

theta_out: A vector of optimized score index value.

Hval: The negative log likelihood criterion.

DHval: The first derivative of the negative likelihood.

D2Hval: The second derivative of the negative likelihood.

iter: The number iterations used.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

theta.distn, Hfun, DHfun, theta2arclen, testscore, scoreDensity

---

thetasearch *Ensure that estimated score index is global*

Description

Multiple minima are found quite often in the data fitting function that is minimized using function thetafun, and in roughly 10 percent of the estimates there is a minimum that it lower than that detected. The function searches a mesh of 101 points for minima, computes the fitting function at the minima, and assigns the location of the global minimum as the replacement theta if the location differs by more than 0.5 from the value identified by thetafun. The function values and their first two derivatives are also replaced.

Usage

```
thetasearch(WfdList, U, theta, Hval, DHval, D2Hval, thetaind=1:N)
```

Arguments

WfdList	A list vector containing specifications of surprisal curves for each item.
U	An N by n matrix containing indices of chosen items for each test taker.
theta	A vector containing all the score index values.
Hval	A vector containing the N function values.
DHval	A vector containing the N first derivative values.
D2Hval	A vector containing the N second derivative values.
thetaind	A vector containing indices of values to be processed.

Value

A named list object containing objects produced from analyzing the simulations, one set for each simulation:

theta: A vector containing all the score index values including those that are altered.

Hval: A vector containing the N function values included those that are altered.

DHval: A vector containing the N first derivative values included those that are altered.

D2Hval: A vector containing the N second derivative values included those that are altered.

changeindex: Indices of the theta values that are altered

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

thetafun

Examples

```
#  Search for values of theta that are not at the global minimum of the
#  fitting function and replace them as well as their function and
#  derivative values associated with the fine grid value nearest the
#  the global minimum.
WfdList <- Quantshort_parList$WfdList
theta   <- Quantshort_parList$theta
Hval    <- Quantshort_parList$Hval
DHval   <- Quantshort_parList$DHval
D2Hval  <- Quantshort_parList$D2Hval
U       <- Quantshort_dataList$U
Result  <- thetasearch(WfdList, U, theta, Hval, DHval, D2Hval, thetaind=1:1000)
changeindex <- Result$changeindex
print(paste("Number changed =",length(changeindex)))
change  <- theta[changeindex] - Result$theta[changeindex]
oldpar <- par(no.readonly=TRUE)
hist(change,101)
par(oldpar)
```

Ushort *Test data for 24 math calculation questions from the SweSAT Quantitative data.*

Description

These data are for a randomly selected subset of 1000 examinees.

Usage

```
Ushort
```

Format

A .txt file with 1000 rows each containing 24 integers in string format. The integers indicate which answer was chosen for each question by the examinee associated with the row.

Details

The code above inputs the contents of the file using the scan function into a single character vector of length 24,000. The strings are converted to integers by the third command, and the integer vector is then reformatted into a 1000 by 24 matrix containing choice indices. The subsequent analysis of these data is described in detail in the vignette `SweSATQuanitativeShort`.

`Usimulate` *Simulate a test or scale data matrix.*

Description

Used in `dataSimulation`, this function sets up an N by n matrix of index values that specify the index of the option chosen by an examinee or respondent for a specific question.

Usage

```
Usimulate(theta.pop, WfdList)
```

Arguments

`theta.pop`	A vector containing population score index values at which data are to be simulated.
`WfdList`	A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of `WfdList` is a named list containing information computed during the analysis. These named lists contain these objects: **Wfd:** A functional data object containing the `M` surprisal curves f. or a question. **M:** The number of options. **Pbin:** A matrix containing proportions at each bin. **Wbin:** A matrix containing surprisal values at each bin. **Pmatfine:** A matrix of probabilities over a fine mesh. **Wmatfine:** A matrix of surprisal values over a fine mesh. **DWmatfine:** A matrix of the values of the first derivative of surprisal curves over fine mesh. **D2Wmatfine:** A matrix of the values of the second derivative of surprisal curves over fine mesh.

Details

For each question and each examinee a vector of random multinomial integer values is generated using the probability transforms of the surprisal curves and the examinee's score index value.

Value

An N by n matrix of integer index values.

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

`Wbinsmth` *Estimate the option probability and surprisal curves.*

Description

The surprisal curves for each item are fit to the surprisal transforms of choice probabilities for each of a set of bins of current performance values `theta`. The error sums of squares are minimized by the surprisal optimization `smooth.surp` in the `fda` package. The output is a list vector of length `n` containing the functional data objects defining the curves.

Usage

```
Wbinsmth(theta, dataList, WfdList=dataList$WfdList,
                   thetaQnt=seq(0,100, len=2*nbin+1), wtvec=matrix(1,n,1),
                   iterlim=20, conv=1e-4, dbglev=0)
```

Arguments

`theta` A vector of length N containing current values of score index percentile values.

`dataList` A list that contains the objects needed to analyse the test or rating scale with the following fields:

U: A matrix of response data with N rows and n columns where N is the number of examinees or respondents and n is the number of items. Entries in the matrices are the indices of the options chosen. Column i of U is expected to contain only the integers `1,...,noption`.

optList: A list vector containing the numerical score values assigned to the options for this question.

key: If the data are from a test of the multiple choices type where the right answer is scored 1 and the wrong answers 0, this is a numeric vector of length n containing the indices the right answers. Otherwise, it is NULL.

WfdPar: An fdPar object for the defining the surprisal curves.

noption: A numeric vector of length n containing the numbers of options for each item.

nbin: The number of bins for binning the data.

scrrng: A vector of length 2 containing the limits of observed sum scores.

scrfine: A fine mesh of test score values for plotting.

scrvec: A vector of length N containing the examinee or respondent sum scores.

itemvec: A vector of length n containing the question or item sum scores.

percntrnk: A vector length N containing the sum score percentile ranks.

thetaQnt: A numeric vector of length 2*nbin + 1 containing the bin boundaries alternating with the bin centers. These are initially defined as `seq(0,100,len=2*nbin+1)`.

Wdim: The total dimension of the surprisal scores.

PcntMarkers: The marker percentages for plotting: 5, 25, 50, 75 and 95.

`WfdList` A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of `WfdList` is a named list containing information computed during the analysis. These named lists contain these objects:

Wfd: A functional data object containing the `M` surprisal curves for a question.

M: The number of options.

Pbin: A matrix containing proportions at each bin.

Wbin: A matrix containing surprisal values at each bin.

indfine: A vector of length 101 containing equally spaced plotting points.

Pmatfine: A matrix of probabilities over a fine mesh.

Wmatfine: A matrix of surprisal values over a fine mesh.

DWmatfine: A matrix of the values of the first derivative of surprisal curves over fine mesh.

D2Wmatfine: A matrix of the values of the second derivative of surprisal curves over fine mesh.

By default `Wbinsmth` uses the `WfdList` member of `dataList`, which is generated by the initialization with `Wbinsmth.init` within `make.dataList`. Usually this is what is required. However, `WfdList` may occasionally need to be entered from another source, and so included here as a separate argument.

`thetaQnt` A vector of length 2*n+1 containing the sequence of bin boundary and bin centre values.

`wtvec` A vector of length `n` of weights on observations. Defaults to all ones.

`iterlim` The maximum number of iterations used in optimizing surprisal curves. Defaults to 20.

`conv` Convergence tolerance. Defaults to 0.0001.

`dbglev` Level of output within `Wbinsmth`. If 0, no output, if 1 the error sum of squares and slope on each iterations, and if 2 or higher, results for each line search iteration with function `lnsrch`.

Details

The function first bins the data in order to achieve rapid estimation of the option surprisal curves. The argument `thetaQnt` contains the sequence of bin boundaries separated by the bin centers, so that it is of length `2*nbin + 1` where `nbin` is the number of bins. These bin values are distributed over the percentile interval [0,100] so that the lowest boundary is 0 and highest 100. Prior to the call to `Wbinsmth` these boundaries are computed so that the numbers of values of `theta` falling in the bins are roughly equal. It is important that the number of bins be chosen so that the bins contain at least about 25 values.

After the values of `theta` are binned, the proportions that the bins are chosen for each question and each option are computed. Proportions of zero are given NA values.

The positive proportions are then converted to surprisal values where surprisal = -log_M (proportion) where log_M is the logarithm with base M, the number of options associated with a question. Bins with zero proportions are assigned a surprisal that is appropriately large in the sense of being in the range of the larger surprisal values associated with small but positive proportions.

The next step is to fit the surprisal values for each question by a functional data object that is smooth, passes as closely as possible to an option's surprisal values, and has values consistent with being a surprisal value. The function `smooth.surp()` is used for this purpose.

Finally the curves and other results for each question are saved in object `WfdList`, a list vector of length n, and the list vector is returned.

Value

The optimized numbered list object produced by `Wbinsmth`. See above for its description.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

See Also

`ICC.plot`, `thetafun`

Examples

```
#  Example 1.  Display the item probability and surprisal curves for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
#  estimate the initial surprisal curves
theta     <- Quantshort_dataList$percntrnk
thetaQnt  <- Quantshort_parList$thetaQnt
Qvec      <- Quantshort_parList$Qvec
WfdResult <- Wbinsmth(theta, Quantshort_dataList)
WfdList   <- WfdResult$WfdList
#  plot the curves for the first question
binctr    <- Quantshort_parList$binctr
scrfine   <- seq(0,100,len=101)
ICC.plot(scrfine, WfdList, Quantshort_dataList, Qvec, binctr,
         plotindex=1, plotrange=c(0,100))
#  estimate the final surprisal curves
theta    <- Quantshort_parList$theta
thetaQnt <- Quantshort_parList$thetaQnt
Qvec     <- Quantshort_parList$Qvec
WfdList  <- Quantshort_parList$WfdList
#  plot the curves for the first question
ICC.plot(scrfine, WfdList, Quantshort_dataList, Qvec, binctr,
         plotindex=1, plotrange=c(0,100))
#  Example 2.  Display the item probability and surprisal curves for the
# Symptom Distress Scale with 13 items and 473 respondents.
```

```
# Proceed as above changing "Quant" to "SDS"
```

`Wpca.plot` *Plot the test information or scale curve in either two or three dimensions.*

Description

A test or scale analysis produces a space curve that varies with in the space of possible option curves of dimension `Wdim`. Fortunately, it is usual that most of the shape variation in the curve is within only two or three dimensions, and these can be fixed by using functional principal components analysis.

Usage

```
Wpca.plot(arclength, WfdList, Wdim, nharm=2, rotate=TRUE, dodge=1.003,
titlestr=NULL, Display=TRUE)
```

Arguments

`arclength`	The total length of the test information or scale curve as computed by function `theta2arclen`.
`WfdList`	A numbered list object produced by a TestGardener analysis of a test. Its length is equal to the number of items in the test or questions in the scale. Each member of `WfdList` is a named list containing information computed during the analysis. These named lists contain these objects: **Wfd:** A functional data object containing the `M` surprisal curves for a question. **M:** The number of options. **Pbin:** A matrix containing proportions at each bin. **Wbin:** A matrix containing surprisal values at each bin. **Pmatfine:** A matrix of probabilities over a fine mesh. **Wmatfine:** A matrix of surprisal values over a fine mesh. **DWmatfine:** A matrix of the values of the first derivative of surprisal curves over fine mesh. **D2Wmatfine:** A matrix of the values of the second derivative of surprisal curves over fine mesh.
`Wdim`	The total number of options in the test or scale.
`nharm`	The number of principal components of the test information or scale curve to be used to display the curve. Must be either 2 or 3.
`rotate`	If true, rotate principal components of the test information or scale curve to be used to display the curve to VARIMAX orientation.
`dodge`	A constant greater than 1 required by ggplot2. Defaults to 1.003.
`titlestr`	A string for the title of the plot. Defaults to NULL.
`Display`	A logical value that is TRUE if the test information manifold is to be displayed.

Value

A named list with these members:

pcaplot	If two dimensions or harmonics are specified, this is a gg or ggplot object that can be displayed using the print command. If three dimensions are specified, this is NULL.
harmvarmxfd	Functional data objects for the principal components of the curve shape.
varpropvarmx	Proportions of variance accounted for by the principal components

The principal components are VARIMAX rotated by default. The plot is displayed as a side value even if no output object is specified in the call to the function.

Author(s)

Juan Li and James Ramsay

References

Ramsay, J. O., Li J. and Wiberg, M. (2020) Full information optimal scoring. Journal of Educational and Behavioral Statistics, 45, 297-315.

Ramsay, J. O., Li J. and Wiberg, M. (2020) Better rating scale scores with information-based psychometrics. Psych, 2, 347-360.

http://testgardener.azurewebsites.net

Examples

```
#  Example 1.  Display the test information curve for the
#  short SweSAT multiple choice test with 24 items and 1000 examinees
#  plot a two-dimension version of manifold curve
WfdList    <- Quantshort_parList$WfdList
theta      <- Quantshort_parList$theta
arclength  <- Quantshort_parList$arclength
oldpar <- par(no.readonly=TRUE)
on.exit(oldpar)
Wpca.plotResults <- Wpca.plot(arclength, WfdList, Quantshort_dataList$Wdim)
varprop <- Wpca.plotResults$varpropvarmx
print("Proportions of variance accounted for and their sum:")
print(round(c(varprop,sum(varprop)),3))
#  plot a three-dimension version of manifold curve
WfdList    <- Quantshort_parList$WfdList
theta      <- Quantshort_parList$theta
arclength  <- Quantshort_parList$arclength
Wpca.plotResults <- Wpca.plot(arclength, WfdList, Quantshort_dataList$Wdim, nharm=3)
varprop <- Wpca.plotResults$varpropvarmx
print("Proportions of variance accounted for and their sum:")
print(round(c(varprop,sum(varprop)),3))
#  Example 2.  Display the test information curve for the
# Symptom Distress Scale with 13 items and 473 respondents.
# Proceed as above changing "Quant" to "SDS"
```

Index