

# Package ‘LatentREGpp’

November 28, 2016

**Type** Package

**Title** Item Response Theory Implemented in R and Cpp

**Version** 0.2.2

**Maintainer** Camilo Antonio Suarez Bolanos <caasuarezbo@unal.edu.co>

**Description** Provides a C++ implementation of the Multidimensional Item Response Theory (MIRT) capable of performing parameter and traits estimations. It also provides a list of options to perform an optimal analysis and obtain useful information about the resulting model. This package is a work of SICS Research Group, Universidad Nacional de Colombia.

**License** MIT + file LICENSE

**LazyData** TRUE

**Imports** Rcpp (>= 0.12.6), randtoolbox, fastGHQuad, RSpectra,  
FactoMineR, MASS, ade4, sirt, paran, statmod, optimx, mvtnorm,  
numDeriv, stats, utils

**LinkingTo** Rcpp, BH

**SystemRequirements** c++11, OpenMP

**RoxygenNote** 5.0.1

**NeedsCompilation** yes

**Author** Milder Hernandez Cagua [aut],  
Jhonatan Javier Guzman del Rio [aut],  
Camilo Antonio Suarez Bolanos [cre],  
Alvaro Mauricio Montenegro [aut],  
Sergio Paez Moncaleano [aut],  
Juan David Cortes Castillo [aut],  
Campo Elias Pardo Turriago [aut],  
Luisa Fernanda Parra Arboleda [aut],  
Emilio Pablo Berdugo Camacho [aut],  
SICS Research Group [cph]

**Repository** CRAN

**Date/Publication** 2016-11-28 22:04:18

**R topics documented:**

|                                   |           |
|-----------------------------------|-----------|
| alpha_cronbach . . . . .          | 2         |
| alpha_curve . . . . .             | 3         |
| an_parallel . . . . .             | 4         |
| biserial_cor . . . . .            | 5         |
| eap_gpc . . . . .                 | 6         |
| envelope_itemf . . . . .          | 7         |
| estim_gpc . . . . .               | 8         |
| findim_hcpc . . . . .             | 8         |
| gutt . . . . .                    | 9         |
| itemfit . . . . .                 | 10        |
| itemfit_bayesian . . . . .        | 12        |
| LatentREGpp-Description . . . . . | 13        |
| map_gpc . . . . .                 | 14        |
| orlando_itemf . . . . .           | 15        |
| personfit . . . . .               | 16        |
| phi . . . . .                     | 18        |
| polyserial_cor . . . . .          | 19        |
| quadpoints . . . . .              | 20        |
| simulate_dichotomous . . . . .    | 20        |
| simulate_polytomous . . . . .     | 21        |
| x2_itemf . . . . .                | 22        |
| Yule . . . . .                    | 22        |
| z3_itemf . . . . .                | 23        |
| z3_personf . . . . .              | 24        |
| <b>Index</b>                      | <b>25</b> |

---

|                |                         |
|----------------|-------------------------|
| alpha_cronbach | <i>Cronbach's alpha</i> |
|----------------|-------------------------|

---

**Description**

Cronbach's alpha measures how correlated are the items in a test. Taken from ltm::cronbach.alpha

**Usage**

```
alpha_cronbach(data, standardized, CI, probs, B, na.rm)
```

**Arguments**

|              |   |
|--------------|---|
| data         | a matrix or a Dataframe that holds the test response data |
| standardized | Standarize data   |
| CI           | boolean   |
| probs        | Vector of probabilities                                   |
| B            | Vector Size   |
| na.rm        | Remove Na values  |

**Details**

the coefficient is calculated

$$\alpha = (n/n - 1) * (1 - (\sum V_i/V_t))$$

where

$$V_t$$

is the variance of test scores and

$$V_i$$

is the variance of item scores. It is desirable that the items are closely interrelated (coefficient near 1). This function was extracted from multilevel\_2.5 package.

**Value**

Cronbach's alpha for the test and the number of individuals of test.

**References**

Cronbach L. J. (1951) Coefficient Alpha and the internal structure of tests. Psychometrika, 16,297-334

---

alpha\_curve

*Cronbach-Mesbah Curve*

---

**Description**

Assess the unidimensionality of a set of items from alpha coefficient.

**Usage**

```
alpha_curve(data)
```

**Arguments**

data            a Dataframe that holds the test response data

**Details**

Constructs the curve by iteration:

1. Compute alpha using all items.
2. Remove one item from the scale. The removed item is the one which leaves the scale with its maximum alpha value. If we remove a bad item, the alpha coefficient will increase. Whereas if we remove a good item, the alpha must decrease.
3. Repeat until only two items remain.

This function was extracted from CMC\_1.0 package.

**Value**

- The number of items used to calculate the coefficient.
- The maximum value of the alpha coefficient calculated at each step.
- The item removed at each step.
- The Cronbach-Mesbah curve plot.

**References**

- Cameletti, M. & Caviezel, V. (2010). Checking the unidimensionality in R using the Cronbach-Mesbah curve.
- Mesbah, M. (2010). Statistical quality of life. In "Method and Applications of Statistics in the Life and Health Sciences", N. BalakrishnanEd., Wiley, pp. 839-864.

---

 an\_parallel

*Parallel Analysis*


---

**Description**

Performs Horn's parallel analysis for a principal component.

**Usage**

```
an_parallel(x = NA, iterations = 0, centile = 0,
  seed = 0, mat = NA, n = NA)
```

**Arguments**

- |            |   |
|------------|---|
| x          | a matrix or a dataframe that holds the test response data.                                      |
| iterations | how many iterations are performed during the analysis. Amount of random datasets produced.      |
| centile    | a number between 1 and 99 indicating the centile used in estimating bias.                       |
| seed       | seed used for the random data generation.   |
| mat        | a supplied correlation matrix to use instead of x. If used, then n must be also supplied.       |
| n          | the number of observations. Required when a correlation matrix is supplied with the mat option. |

**Details**

Implementation of Horn's (1965) technique for evaluating the components retained in a principle component analysis (PCA). This procedure is an adaptation of the paran function in the Paran Package.

**Value**

Retained Components a scalar integer representing the number of retained components.

Adjusted Eigenvalues a vector of the estimated adjusted eigenvalues.

Unadjusted Eigenvalues a vector of the eigenvalues of the observed data from an unrotated principal component analysis.

Bias a vector of the estimated bias of the unadjusted eigenvalues

**References**

John L. Horn (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, Volume 30, Number 2, Page 179.

Dinno A. 2009. Exploring the Sensitivity of Horn's Parallel Analysis to the Distributional Form of Simulated Data. *Multivariate Behavioral Research*. 44(3): 362-388

**Examples**

```
## Not run:
data <- simulate_dichotomous(size.cluster = c(10), sample.size=1000)
an_parallel(data$data, iterations = 100, centile = 99, seed = 12)

## End(Not run)
```

---

biserial\_cor

*Biserial Correlation*


---

**Description**

Point-Biserial correlation coefficient is a correlation coefficient used when one variable is continuous and the other variable is dichotomous. Taken from `ltm::biserial_cor`

**Usage**

```
biserial_cor(x,y, use = c("all.obs", "complete.obs"), level = 1)
```

**Arguments**

|       |   |
|-------|---|
| x     | a numeric vector representing the continuous variable.  |
| y     | a numeric vector representing the dichotomous variable. |
| use   | is a option for the use of missing values.              |
| level | which level of y to use.                                |

**Details**

It is calculated by applying the Pearson correlation coefficient to the case where one of the variables has dichotomous nature.

It is calculated as

$$r_{xy} = (\bar{x}_p - \bar{x}_q / S_x) * \sqrt{pq}$$

Where p is the proportion of subjects with one of the two possible values of the variable Y, q is the proportion of subjects with the other possible value,

$$\bar{x}_p$$

and

$$\bar{x}_q$$

is the average X subjects whose proportion is p and q respectively, and

$$S_x$$

is the standard deviation of all subjects X. This function was adapted from ltm\_1.0 package.

**Value**

The value of the point-biserial correlation.

**References**

U.Olsson, F.Dragow, and N.Dorans (1982). The polyserial correlation coefficient. *Psychometrika*, 47:337-347.

Cox. N.R. (1974). Estimation of the Correlation between a Continuous and a Discrete Variable. *Biometrics*, 30:171-178.

**Examples**

```
## Not run:
data <- simulate_dichotomous(size.cluster = c(10), sample.size=1000)
biserial_cor(rowSums(data$data), data$data[,1])

## End(Not run)
```

---

eap\_gpc

*Latent Trait Estimation with EAP for generalized partial credit model (gpc) model*

---

**Description**

Estimates latent traits for generalized partial credit model (gpc) model. Just for unidimensional models.

**Usage**

```
eap_gpc(estim)
```

**Arguments**

`estim`                    Output object from the `estim_gpc` function.

**Examples**

```
## Not run:
data = simulate_polytomous()$data
estim=estim_gpc(datos = datos)
eap_gpc = eap_gpc(estim)
plot(density(eap_gpc))

## End(Not run)
```

---

envelope\_itemf                    *Confidence Envelops for an item*

---

**Description**

Graphs confidence bands of an item, to evaluate the goodness of fit of the model.

**Usage**

```
envelope_itemf(item, numboot = 100, alpha = 0.05, item.fit, data,
  seed = 5000L)
```

**Arguments**

`item`                    a number indicating the item to be evaluated.  
`numboot`                number of iterations bootstrap, used to plot the envelopes. By default 100.  
`alpha`                    level of significance to plot the envelopes. By default 0.05.  
`item.fit`                object `LatentREGpp::itemfit()` type.  
`data`                    a dataframe or a matrix with the test data.  
`seed`                    the seed to fix the random sample. By default 500L.

**Value**

plot with the envelopes and the characteristic curve of the item.

**References**

David Thissen, Howard Wainer D. (1990). Confidence Envelopes for Item Response Theory. *Journal of Educational Statistics, Vol 15, No 2*, 113-128.

**See Also**

[orlando\\_itemf](#), [z3\\_itemf](#)

---

estim\_gpc

*Estimation gpc model*

---

**Description**

Estimates the test parameters according to the generalized partial credit model (gpc) model. This model is unidimensional only.

**Usage**

```
estim_gpc(data)
```

**Arguments**

data                    The matrix containing the answers of tested individuals

**Examples**

```
## Not run:
data = simulate_polytomous$data
estim=estim_gpc(data = datos)

## End(Not run)
```

---

findim\_hcpc

*Discovering Dimension*

---

**Description**

Finds the test dimension for a dichotomous logistic model by using Loadings and Factorial Analysis, according to the Paez & Montenegro methodology.

**Usage**

```
findim_hcpc(data, verbose, probit)
```

**Arguments**

data                    A binary matrix or dataframe that holds the response data with N individuals (columns) and P items (rows).

verbose                a boolean, if TRUE all procedures are described in console. The adjusted eigen values are always shown.

probit                 a boolean, if FALSE correction of probit to logit model is made (multiplying by 1.702 de discrimination vectors).



**Details**

Implementation of a technique to evaluate the number of latent constructs presented by data.

**Value**

L\_data\_Clust: list of matrices formed by clusters, each matrix has an associated set of binary vectors of items.

M\_data: matrix of the join of objects in L\_data\_Clust.

dim\_new: numeric, Dimension found by the algorithm.

dim\_old: numeric, Initial Dimension of the algorithm (set by parallel analysis).

sc: vector, The size of each cluster of L\_data\_Clust.

**References**

Paez S, Montenegro A. and Pardo C. (2017). Principles and Methodology to Find Dimension on Latent Regression Models. Novel based approach. British Journal of Mathematical and Statistical Psychology. (Submitted)

John L. Horn (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, Volume 30, Number 2, Page 179.

Lebart L, Morineau A, Piron M (1997). Statistique Exploratoire Multidimensionnelle. Dunod.

Reckase M (2009). Multidimensional item response theory. Springer.

**See Also**

[itemfit](#)

**Examples**

```
## Not run:
file = paste(system.file(package = "LatentREGpp"), "/dataset/TUN.txt", sep = "")
TUN = data.matrix(read.table(file))
findim_hcpc(TUN)
findim_hcpc(TUN, verbose = T)

## End(Not run)
```

---

gutt

*Guttman's Lambda*


---

**Description**

Six Lower limits of reliability coefficients are presented.

**Usage**

```
gutt(data)
```

**Arguments**

data                    a matrix or a Dataframe that holds the test response data

**Details**

Let

$$S_j^2$$

the variances over persons of the  $n$  items in the test, and  $S_t^2$  the variance over persons of the sum of the items. The first estimate  $\lambda_1$  can be computed from  $L_1 = 1 - (\sum s_j^2 / S_t^2)$ . Let  $C_2$  the sum of squares of the covariances between items, therefore is the sum of  $n(n-1)/2$  terms. The bound  $\lambda_2$  is computed by  $L_2 = L_1 + (\sqrt{n/n-1} C_2 / S_t^2)$ . The third lower bound  $\lambda_3$  is a modification of  $\lambda_1$ , it is computed from the  $L_3 = n/(n-1)L_1$ . Fourth lower bound  $\lambda_4$  has been interpreted as the greatest split half reliability, and requires that the test be scored as two halves. It is calculated from

$$L_4 = 2(1 - (s_a^2 + s_b^2) / s_t^2)$$

where  $S_a^2$  and  $S_b^2$  are the respective variances of the two parts for the single trial. For the fifth lower bound  $\lambda_5$ , let  $C_{2j}$  be the sum of the squares of the covariances of item  $j$  with the remaining  $n-1$  items, and let  $\bar{C}_2$  be the largest of the

$$C_{2j}$$

. Then the coefficient can be computed from  $L_5 = L_1 + (2\sqrt{\bar{C}_2}) / S_t^2$ . The final bound is based on multiple correlation, let  $e_j^2$  be the variance of the errors of estimate of item  $j$  from its linear multiple regression on the remaining  $n-1$  items. Then  $\lambda_6$  can be computed from  $L_6 = 1 - (\sum e_j^2) / S_t^2$

**Value**

The six coefficients Guttman for the test.

**References**

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255-282.

---

itemfit

*Parameter estimation of a test*

---

**Description**

Estimates the test parameters according to the Multidimensional Item Response Theory

**Usage**

```
itemfit(data, dim, model = "2PL", EMepsilon = 1e-04, clusters = NULL,
quad_tech = NULL, quad_points = NULL,
individual_weights = as.numeric(c()), initial_values = NULL, SD = FALSE,
verbose = TRUE, save_time = TRUE)
```

**Arguments**

|                    |  |
|--------------------|--|
| data               | The matrix containing the answers of tested individuals  |
| dim                | The dimensionality of the test   |
| model              | "1PL", "2PL" or "3PL"  |
| EMepsilon          | Convergence value to determine the accuracy of the test  |
| clusters           | A vector with cluster per dimension  |
| quad_tech          | A string with technique. "Gaussian" for Gaussian quadrature or "QMCEM" for Quasi-Monte Carlo quadrature        |
| quad_points        | Amount of quadrature points. If quadratura_technique is "Gaussian". It can be NULL                             |
| individual_weights | A vector with Weights of the quadrature points.  |
| initial_values     | A matrix with initial values for estimation process. Be sure about dimension, model and consistency with data. |
| SD                 | calculate for standar desviation for items   |
| verbose            | True for get information about estimation process in runtime. False in otherwise.                              |
| save_time          | True for save estimation time. False otherwise.  |

**Models**

LatentREGpp has different models to fit likelihood value according parameters to estimate.

**3PL** General. Probability is given by

$$P_{ij} = c_j + \frac{1 - c_j}{1 + \exp(-\eta_{ij})}$$

Where  $i$  references individual and  $j$  references the item;  $c$  is a value for guessing parameter between 0 and 1.  $i$  index is referenced by number of examinees or individuals and  $j$  index is referenced by items in test.  $\eta$  is

$$\eta_{ij} = \mathbf{a}_j^t \theta_i + d_j$$

In unidimensional an  $a$  is scalar, in multidimensional an  $\mathbf{a}$  is vector. For 1PL model  $a$  has value 1

**2PL**  $c = 0$

**1PL**  $c = 0$   $\mathbf{a}$  vector has a value 1 for each element

**Examples**

```
## Not run:
#Example 1

dir = normalizePath(system.file(package="LatentREGpp"),winslash = "/")
folder = "/dataset/1D/dicho/"
file = "1000x50-1.csv"
data_dir = paste(c(dir, folder, file), collapse = "")
data = read.table(file = data_dir, sep = ";")
```

```

est <- itemfit(data = data, dim = 1)

#Example 2

#Dichotomous and multidimensional data
dir = normalizePath(system.file(package="LatentREGpp"),winslash = "/")
folder = "/dataset/3D/dicho/"
file = "1000x55-1.csv"
data_dir = paste(c(dir, folder, file), collapse = "")
data = read.table(file = data_dir, sep = ";")
clust <- c(20,20,15)
st <- itemfit(data = data, model = "2PL",dim = 3,
              EMepsilon = 1e-03, clusters = clust, quad_tech = "Gaussian")

## End(Not run)

```

---

itemfit\_bayesian

*Bayesian parameter estimation of a test*


---

## Description

Estimates the test parameters according to the Multidimensional Item Response Theory with bayesian adjust for dichotomous data

## Usage

```

itemfit_bayesian(data, dim, model = "2PL", EMepsilon = 1e-04,
                 clusters = NULL, quad_tech = NULL, quad_points = NULL,
                 individual_weights = as.numeric(c()), initial_values = NULL,
                 noguessing = TRUE, verbose = TRUE, save_time = TRUE)

```

## Arguments

|                    |   |
|--------------------|---|
| data               | The matrix containing the answers of tested individuals   |
| dim                | The dimensionality of the test  |
| model              | "1PL", "2PL" or "3PL"   |
| EMepsilon          | Convergence value to determine the accuracy of the test   |
| clusters           | A vector with cluster per dimension   |
| quad_tech          | A string with technique. "Gaussian" for Gaussian quadrature or "QMCEM" for Quasi-Monte Carlo quadrature. If NULL it's selected according to the model's dimension (QMCEM if dim>3). |
| quad_points        | Amount of quadrature points. If quadratura_technique is "Gaussian". It can be NULL  |
| individual_weights | A vector with Weights of the quadrature points.   |
| initial_values     | A matrix with initial values for estimation process. Be sure about dimension, model and consistency with data.  |

|            |   |
|------------|---|
| noguessing | In 3PL model and dimension is greater than 1, If true, guessing parameter will not be estimated in zeta vector. Instead c value will have a default initial value. Otherwise guessing parameter will be estimated with zeta vector. |
| verbose    | True for get information about estimation process in runtime. False in otherwise.   |
| save_time  | True for save estimation time. False otherwise.   |

## Model

Bayesian model is based in itemfit models. It has a  $Q_i$  function to optimize according parameters like in itemfit. However this model is given by:

$$Q_i = N * \log(P_{\zeta_i}(\zeta_i)) + \hat{Q}_i$$

Where i index is referenced for items in test.

Then, log posterior is given by:

$$\log(P_{\zeta_i}(\zeta_i)) = -\frac{N}{2} \left( \frac{(a_{1i} - \mu_{a1i})^2}{\sigma_{1i}^2} + \dots + \frac{(a_{Di} - \mu_{aDi})^2}{\sigma_{Di}^2} + \frac{(d_i - \mu_{di})^2}{\sigma_{di}^2} + \frac{(c_i - \mu_{ci})^2}{\sigma_{ci}^2} \right)$$

Where a,d and c are parameters, D is the dimension of test. You can give the  $\mu$  values for each parameters through initial values matrix. In otherwise  $\mu$  will have default initial values value  $\sigma^2$  values are constant  $\sigma_a^2 = 0.64$ ,  $\sigma_d^2 = 4$ ,  $\sigma_c^2 = 0.009$

---

LatentREGpp-Description

*LatentREGpp : Item Response Theory Implemented in R and Cpp*

---

## Description

Provides a C++ implementation of the Multidimensional Item Response Theory (MIRT) capable of performing parameter and traits estimations. It also provides a list of options to perform an optimal analysis and obtain useful information about the resulting model.

This package is a work of SICS Research Group, Universidad Nacional de Colombia.

## Details

Package: LatentREGpp  
 Type: Package  
 Version: 0.2.2  
 Date: 2016-11-28  
 License: MIT + file LICENSE

## Getting Started

Get started with the LatentREGpp package by browsing the index of this documentation if you need help the vignettes should be helpful.

The LatentREGpp package allows you to use the LatentREGpp methodology for simulating, analyzing and scoring tests

You can browse the package vignettes to get started.

## Acknowledgment

This work was Supported by Colciencias Research Grant 0039-2013 and SICS Research Group, Universidad Nacional de Colombia.

## Author(s)

Milder Hernandez Cagua <milderhc@gmail.com>

Jhonatan Javier Guzman del Rio <jhgzmanri@unal.edu.co>

Camilo Antonio Suarez Bolanos <caasuaresbo@unal.edu.co>

Alvaro Mauricio Montenegro <ammontenegrod@unal.edu.co>

Sergio Paez Moncaleano <spaezm@unal.edu.co>

Juan David Cortes Castillo <jdcortesc@unal.edu.co>

Campo Elias Pardo Turriago <cepardot@unal.edu.co>

Luisa Fernanda Parra Arboleda <lparraar@unal.edu.co>

Emilio Pablo Berdugo Camacho <epberdugoc@unal.edu.co>

---

map\_gpc

*Latent Trait Estimation with MAP for generalised partial credit model (gpc)*

---

## Description

Estimates latent traits for generalised partial credit model (gpc) with MAP data matrix for estimation. Just for unidimensional models.

## Usage

```
map_gpc(estim)
```

## Arguments

estim            Output object from the estim\_gpc function.

**Examples**

```
## Not run:
data = simulate_polytomous()$data
estim=estim_gpc(datos = datos)
map_gpc = map_gpc(estim)

## End(Not run)
```

---

|               |                            |
|---------------|----------------------------|
| orlando_itemf | <i>Orlando's statistic</i> |
|---------------|----------------------------|

---

**Description**

Calculate the values of the statistics  $S_{x2}$  from Maria Orlando and David Thisen (2000).

**Usage**

```
orlando_itemf(patterns, G, zeta, model)
```

**Arguments**

|          |   |
|----------|---|
| patterns | list of patterns response (patterns\$patterns), the frequency of each pattern (patterns\$freq) and the latent traits (patterns\$latent_traits). |
| G        | number of quadrature points.  |
| zeta     | matrix of estimations of the parameters of the items (alphas, d's, guessing).   |
| model    | type of model ( "1PL", "2PL", "3PL" ).  |

**Value**

Orlando's statistic, degrees of freedom and p-value for each item.

**References**

Orlando, M. & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64.

**See Also**

[z3\\_itemf](#)

---

 personfit

*Latent traits estimation*


---

### Description

Estimates the latent traits by using either the Expected A Posteriori (EAP) or Mode A Posteriori (MAP) method. A Normal distribution with mean vector zero and covariance matrix the identity is assumed. Quasi-Monte Carlo quadrature is suggested when the data dimension is large ( $> 3$ ).

### Usage

```
personfit(data, dim, model = "2PL", zetas, quad_tech = "Gaussian",
  quad_points = NULL, init_traits = NULL, method = "MAP",
  by_individuals = TRUE, verbose = FALSE)
```

### Arguments

|                |   |
|----------------|---|
| data           | The matrix containing the answers of tested individuals   |
| dim            | The dimensionality of the test  |
| model          | "1PL", "2PL" or "3PL"   |
| zetas          | The item parameters. A matrix of dim (num of items * num of parameters from item that has the greater number of categories) where each row is a vector of the form:<br>$(\mathbf{a}_i, \gamma_{i1}, \gamma_{i2}, \dots, \gamma_{im_i}, NA, NA, \dots, NA, c_i)$ according to the notation in the section "Notation". The function "LatentREGpp::itemfit()" returns the zetas with this structure. |
| quad_tech      | A string with technique. "Gaussian" for Gaussian quadrature or "QMCEM" for Quasi-Monte Carlo quadrature   |
| quad_points    | Amount of quadrature points by dimension. Default NULL  |
| init_traits    | Initial values by pattern or by individual as the case may be. Default NULL.  |
| method         | "EAP" or "MAP". "MAP" by default.   |
| by_individuals | if True, return latent trait by individual, otherwise by response pattern. True by default.   |
| verbose        | True for get information about estimation process in runtime. False in otherwise.   |

### Value

depends on value of by\_individuals argument.

If by\_individuals=TRUE. Returns a matrix with latent trait for each individual

If by\_individuals=FALSE. Returns a list of length 3 with the latent traits for each pattern, the response patterns and the frequency of each pattern



### Methods to estimate the Latent Traits

**EAP** In general the EAP Method is based on the next expression.

$$\frac{\int \theta_l p(U_l = u_l/\theta_l, \zeta) p(\theta_l/\eta) \partial \theta_l}{\int p(U_l = u_l/\theta_l, \zeta) p(\theta_l/\eta) \partial \theta_l}$$

where:

$\theta_l$  is the latent trait associated with pattern l.

$U_l$  refers to response pattern l

$\zeta$  are items parameters

$\eta$  are the hiperparameters from the prior distribution to the traits.

**MAP** The method consists of maximize the following expression regard to  $\theta_l$

$$\frac{p(U_l = u_l/\theta_l, \zeta) p(\theta_l/\eta)}{\int p(U_l = u_l/\theta_l, \zeta) p(\theta_l/\eta) \partial \theta_l}$$

### Notation

In the Polytomous Multidimensional case, the probability that an examinee with latent trait vector  $\theta_l$  responses categorie k to item i is,

$$P(U_{li} = k \mid \theta_l, \mathbf{a}_i, \gamma_{ik}) = c_i + (1 - c_i) \Psi(\eta_{lik})$$

where

$$\eta_{lik} = \mathbf{a}_i^t \theta_l + \gamma_{ik}$$

$\mathbf{a}_i = (\mathbf{a}_{i1}, \mathbf{a}_{i2}, \dots, \mathbf{a}_{id})^t$  is a parameter associated with discrimination of item, d is the data dimension

$\theta_l = (\theta_{l1}, \theta_{l2}, \dots, \theta_{ld})^t$  is the latent trait multidimensional associated with pattern l.

$\gamma_{ik}$  is a parameter associated with item and the categorie,  $i = 1, 2, \dots, p$  (number of items)  $k = 1, 2, \dots, m_i$  (number of categories from item)

$c_i$  is the guessing parameter, make sense in the dichotomous case.

### Examples

```
## Not run:
#Example 1

#simulate 10 polyotmous items, the first 5 with 4 response categories and the
#others with 5 response categories, By default the number of individuals is 1000
#, model is "2PL" and the dimation is 1
dats=simulate_polytomous(ncatgs = c(rep(4,10),rep(5,10)),seed_data = 5000L)
#estimate the items parameters
est=itemfit(dats$data,dim = 1,model = "2PL")
#calculates the latent traits estimation.
personfit(dats$data,dim = 1,zetas = est$zetas,method = "MAP")

#Example 2
```

```

#simulate 10 dichotomous items, the trait dimension is 2
#other arguments by default
dats=simulate_dichotomous(dim.data = 2,size.cluster = c(5,5), seed_data = 5000L)
#estimate the items parameters
est=itemfit(dats$data,dim = 2,clusters = c(5,5),model = "2PL")
#calculates the latent traits estimation
personfit(dats$data,dim = 2,zetas = est$zetas,method = "MAP",by_individuals=F)

## End(Not run)

```

---

 phi

*Phi coefficient of correlation*


---

### Description

The phi coefficient is a correlation coefficient applied to dichotomous data. Given a two x two table of counts

|   |   |    |
|---|---|----|
| a | b | R1 |
|---|---|----|

|   |   |    |
|---|---|----|
| c | d | R1 |
|---|---|----|

|  |  |  |
|--|--|--|
|  |  |  |
|--|--|--|

|    |    |   |
|----|----|---|
| C1 | C2 | n |
|----|----|---|

or a vector c(a,b,c,d) of frequencies.

### Usage

```
phi(x)
```

### Arguments

x                    a 1 x 4 vector or a matrix 2 x 2 of frequencies.

### Details

The coefficient phi is calculated from

$$(ad - bc) / \sqrt{p_1 p_2 q_1 q_2}$$

where

$$p_i$$

and

$$q_i$$

are the ratios of the dichotomous variables.

### Value

the value of the phi coefficient correlation.

**References**

Warrens, Matthijs (2008), On Association Coefficients for 2x2 Tables and Properties That Do Not Depend on the Marginal Distributions. *Psychometrika*, 73, 777-789.

Yule, G.U. (1912). On the methods of measuring the association between two attributes. *Journal of the Royal Statistical Society*, 75, 579-652.

---

polyserial\_cor                      *Polyserial correlation*

---

**Description**

Polyserial correlation coefficient is a correlation coefficient used when one variable is continuous and the other variable is dichotomous. Taken and adapted from psych::polyserial

**Usage**

polyserial\_cor(x,y)

**Arguments**

x                      a numeric vector representing the continuous variable.  
y                      a numeric vector representing the dichotomous variable.

**Details**

The coefficient is calculated from

$$\rho = r_{xy} * \sqrt{(n-1)/n} * s_y / \sum \phi(\tau)$$

where

$$r_{xy}$$

is the coefficient of correlation of Pearson coefficient,  $S_y$  is the standard deviation of Y, and

$$\phi(\tau)$$

are the ordinates of the normal curve at the normal equivalent of the cut point boundaries between the item responses. This function was adapted from ltm\_1.0 package.

**Value**

the value of the polyserial correlation.

**References**

U.Olsson, F.Drasgow, and N.Dorans (1982). The polyserial correlation coefficient. *Psychometrika*, 47:337-347.

**Examples**

```
## Not run:
x <- rnorm(100)
y <- sample(1:5,100,replace=TRUE)
cor(x, y)
polyserial_cor(x, y)

## End(Not run)
```

---

|            |                          |
|------------|--------------------------|
| quadpoints | <i>Quadrature points</i> |
|------------|--------------------------|

---

**Description**

Return a list with quadrature points according to dimensionality, technique and number of points.

**Usage**

```
quadpoints(dim, quad_tech = "Gaussian", quad_points = NULL)
```

**Arguments**

|             |  |
|-------------|--|
| dim         | Dimension of the quadrature  |
| quad_tech   | A string to specify the quadrature calculation technique. Use "Gaussian" to use that method, or or "QMCEM" for Quasi-Monte Carlo quadrature.   |
| quad_points | An integer number specifying the amount of quadrature points to use. If NULL, the program will choose the best one. If Quasi-Monte Carlo method is specified, the default value is of 2000 points. |

**Examples**

```
## Not run: qp = quadpoints(dim = 4,quad_tech = "QMCEM",quad_points = 3000)
```

---

|                      |                                    |
|----------------------|------------------------------------|
| simulate_dichotomous | <i>Dichotomous data simulation</i> |
|----------------------|------------------------------------|

---

**Description**

This function a generates dichotomous test

**Usage**

```
simulate_dichotomous(dim.data = 1, sample.size = 1000,
  size.cluster = c(20), seed_data = 5000L, seed_item = 1000,
  model = "2PL")
```

**Arguments**

|              |   |
|--------------|---|
| dim.data     | Data dimension  |
| sample.size  | Size of the population to be simulated  |
| size.cluster | Vector containing the number of items per dimension                               |
| seed_data    | Seed for simulation of the data   |
| seed_item    | Seed for simulation of the item parameters  |
| model        | the model used to generate the test. It can have values of "1PL", "2PL" or "3PL". |

**Value**

list with the simulated data, the information of the item parameters and the individual parameters

**See Also**

[simulate\\_polytomous](#)

**Examples**

```
#Example 1: see by default options
simm=simulate_dichotomous()
```

---

simulate\_polytomous    *Polytomous data simulation*

---

**Description**

This function generates dichotomous test of an specified dimension and size. The items per dimension are specified in the size.cluster vector, and the individuals are specified in sample.size. The amount of categories of each item is determined with ncatgs.

**Usage**

```
simulate_polytomous(dim.data = 1, sample.size = 1000,
  size.cluster = c(20), ncatgs = rep(4, 20), seed_data = 5000L,
  model = "2PL")
```

**Arguments**

|              |   |
|--------------|---|
| dim.data     | Data dimension  |
| sample.size  | Size of the population that will be simulated                                     |
| size.cluster | Vector containing the number of items by dimension                                |
| ncatgs       | Vector containing the number of categories for each item                          |
| seed_data    | Seed for simulation   |
| model        | the model used to generate the test. It can have values of "1PL", "2PL" or "3PL". |

**Value**

list with the simulated data, the information of the item parameters and the individual parameters

**See Also**

[simulate\\_polytomous](#)

---

|          |                        |
|----------|------------------------|
| x2_itemf | <i>Statistical x2.</i> |
|----------|------------------------|

---

**Description**

Calculates the statistical x2.

**Usage**

```
x2_itemf(zetas, patterns, G, FUN)
```

**Arguments**

|          |  |
|----------|--|
| zetas    | matrix of estimations of the parameters of the items (alphas, d's, guessing).                          |
| patterns | list with Patterns, frequencies and traits.  |
| G        | the number of groups, by default 10  |
| FUN      | It is the function with which the expected probability, by default median is calculated in each group. |

---

|      |  |
|------|--|
| Yule | <i>Yule coefficient of correlation</i> |
|------|--|

---

**Description**

The Yule coefficient is a correlation coefficient applied to dichotomous data. Given a two x two table of counts

```
| a | b | R1 |
```

```
| c | d | R1 |
```

```
|---|---|
```

```
|C1 | C2| n |
```

or a vector c(a,b,c,d) of frequencies. Taken from psych::Yule.

**Usage**

```
Yule(x, Y = FALSE)
```

**Arguments**

|   |  |
|---|--|
| x | a 1 x 4 vector or a matrix 2 x 2 of frequencies.         |
| Y | if Y is true return Yule's Y coefficient of colligation. |

**Details**

The coefficient of Yule is calculated from  $(ad - bc)/(ad + bc)$ . This is the number of pairs in agreement (ad) - the number in disagreement (bc) over the total number of paired observations.

**Value**

the value of the Yule Q coefficient.

**References**

Yule, G.U. (1912). On the methods of measuring the association between two attributes. *Journal of the Royal Statistical Society*, 75, 579-652.

Warrens, Matthijs (2008), On Association Coefficients for 2x2 Tables and Properties That Do Not Depend on the Marginal Distributions. *Psychometrika*, 73, 777-789.

---

|          |                             |
|----------|-----------------------------|
| z3_itemf | <i>Z3item fit statistic</i> |
|----------|-----------------------------|

---

**Description**

Calculates the values of statistical Z3 for items.

**Usage**

```
z3_itemf(data, zita, patterns)
```

**Arguments**

|          |   |
|----------|---|
| data     | a data frame or a matrix with the test.   |
| zita     | Matrix which columns are the estimations of the parameters of the items (discrimination,d=discrimination*difficulty, guessing).                 |
| patterns | list of patterns response (patterns\$patterns), the frequency of each pattern (patterns\$freq) and the latent traits (patterns\$latent_traits). |

**References**

Fritz Drasgow, Michael V. Levine and Esther A. Williams (1985). Appropriateness measurement with polychotomous item response models and standardized indices.

**See Also**

[z3\\_personf](#), [orlando\\_itemf](#)

---

|            |                                |
|------------|--------------------------------|
| z3_personf | <i>Z3 Person fit statistic</i> |
|------------|--------------------------------|

---

**Description**

Calculates the values of statistical Z3 for individuals.

**Usage**

```
z3_personf(data, zita, patterns)
```

**Arguments**

|          |   |
|----------|---|
| data     | a data frame or a matrix with the test.   |
| zita     | Matrix whose columns are the estimations of the parameters of the items (discrimination, $d = \text{discrimination} * \text{difficulty}$ , guessing). |
| patterns | list of patterns response (patterns\$patterns), the frequency of each pattern (patterns\$freq) and the latent traits (patterns\$latent_traits).       |

**References**

Fritz Drasgow, Michael V. Levine and Esther A. Williams (1985). Appropriateness measurement with polychotomous item response models and standardized indices.

**See Also**

[z3\\_itemf](#), [orlando\\_itemf](#)



# Index

- \*Topic **IRT**
  - LatentREGpp-Description, 13
- \*Topic **MIRT**
  - LatentREGpp-Description, 13
- \*Topic **Psychometry**
  - LatentREGpp-Description, 13
  
- alpha\_cronbach, 2
- alpha\_curve, 3
- an\_parallel, 4
  
- biserial\_cor, 5
  
- eap\_gpc, 6
- envelope\_itemf, 7
- estim\_gpc, 8
  
- findim\_hcpc, 8
  
- gutt, 9
  
- itemfit, 9, 10
- itemfit\_bayesian, 12
  
- LatentREGpp-Description, 13
- LatentREGpp-Description-package  
(LatentREGpp-Description), 13
  
- map\_gpc, 14
  
- orlando\_itemf, 8, 15, 23, 24
  
- personfit, 16
- phi, 18
- polyserial\_cor, 19
  
- quadpoints, 20
  
- simulate\_dichotomous, 20
- simulate\_polytomous, 21, 21, 22
  
- x2\_itemf, 22
  
- Yule, 22
  
- z3\_itemf, 8, 15, 23, 24
- z3\_personf, 23, 24