

# Package ‘VEMIRT’

December 5, 2024

**Type** Package

**Title** Variational Expectation Maximization for High-Dimensional IRT Models

**Version** 2.6

**Date** 2024-12-05

**Maintainer** Weicong Lyu <weiconglyu@um.edu.mo>

## Description

VEMIRT is created to assist researchers in conducting high-dimensional exploratory and confirmatory multidimensional item response theory (MIRT) analysis and corresponding differential item functioning (DIF) analysis. The core computation engine of VEMIRT is a family of Gaussian Variational EM algorithms that are considerably more efficient than currently available algorithms in other statistical packages, especially when the number of latent factors exceeds four.

**License** GPL-3

**Imports** abind,

GPArotation,

MASS,

Matrix,

mirt,

mvQuad,

mvnfast,

polycor,

psych,

Rcpp,

RcppArmadillo,

testit,

tibble,

torch,

shiny,

shinyjs,

bslib,

data.table,

DT,

callr,

openxlsx

**LinkingTo** Rcpp, RcppArmadillo, RcppEigen

**Encoding** UTF-8

**Depends** R (>= 3.10)

**LazyData** true

**RoxygenNote** 7.3.2

**URL** <https://MAP-LAB-UW.github.io/VEMIRT>, <https://github.com/MAP-LAB-UW/VEMIRT>

**Suggests** knitr,  
rmarkdown

**VignetteBuilder** knitr

## Contents

VEMIRT-package . . . . .	3
C2PL_bs . . . . .	4
C2PL_data . . . . .	5
C2PL_gvem . . . . .	5
C2PL_iw . . . . .	6
C2PL_iw2 . . . . .	8
C3PL_data . . . . .	10
C3PL_sgvem . . . . .	10
coef.vemirt_DIF . . . . .	12
coef.vemirt_DIF_summary . . . . .	12
coef.vemirt_FA . . . . .	13
D2PL_data . . . . .	13
D2PL_em . . . . .	14
D2PL_gvem . . . . .	16
D2PL_lrt . . . . .	18
D2PL_pair_em . . . . .	19
E2PL_data_C1 . . . . .	21
E2PL_data_C2 . . . . .	21
E2PL_gvem_adaptlasso . . . . .	22
E2PL_gvem_lasso . . . . .	24
E2PL_gvem_rot . . . . .	26
E3PL_data_C1 . . . . .	27
E3PL_data_C2 . . . . .	28
E3PL_sgvem_adaptlasso . . . . .	28
E3PL_sgvem_lasso . . . . .	31
E3PL_sgvem_rot . . . . .	33
pa_poly . . . . .	35
print.vemirt_DIF . . . . .	35
print.vemirt_DIF_summary . . . . .	36
print.vemirt_FA . . . . .	36
shinyVEMIRT . . . . .	37
summary.vemirt_DIF . . . . .	37

**Index**

**38**

## Description

VEMIRT is created to assist researchers to conduct exploratory and confirmatory multidimensional item response theory (MIRT) analysis and coresponding item differential functioning (DIF) analysis. The core computation engine of VEMIRT is a family of Gaussian Variational EM algorithms that are considerably more efficient than currently available algorithms in other software packages, especially when the number of latent factors exceeds four.

## Identifying the number of factors

[pa\\_poly](#) identifies the number of factors via parallel analysis.

## Exploratory factor analysis

- [E2PL\\_gvem\\_rot](#) conducts M2PL Analysis with post-hoc rotation (Promax & CF-Quartimax)
- [E2PL\\_gvem\\_lasso](#) conducts M2PL Analysis with Lasso penalty
- [E2PL\\_gvem\\_adaptlasso](#) conducts M2PL Analysis with adaptive Lasso penalty
- [E2PL\\_iw](#) conducts importance sampling to correct bias for M2PL analysis
- [E3PL\\_sgvm\\_rot](#) conducts stochastic GVEM to further improve the computational efficiency for exploratory M3PL analysis
- [E3PL\\_sgvm\\_lasso](#) conducts M3PL Analysis with Lasso penalty
- [E3PL\\_sgvm\\_adaptlasso](#) conducts M3PL Analysis with adaptive Lasso penalty

## Confirmatory factor analysis

- [C2PL\\_gvem](#) conducts GVEM for confirmatory M2PL analysis
- [C2PL\\_bs](#) conducts bootstrap sampling to correct bias and produce standard errors for confirmatory M2PL analysis
- [C2PL\\_iw](#) conducts importance sampling to correct bias for M2PL analysis
- [C2PL\\_iw2](#) conducts IW-GVEM for confirmatory M2PL analysis (alternative implementation to [C2PL\\_iw](#))
- [C3PL\\_sgvm](#) conducts stochastic GVEM for confirmatory M3PL analysis

## Differential item functioning analysis

- [D2PL\\_em](#) conducts DIF analysis for M2PL models using EM algorithms
- [D2PL\\_pair\\_em](#) conducts DIF analysis for 2PL models using EM algorithms with group pairwise truncated  $L_1$  penalty
- [D2PL\\_gvem](#) conducts DIF analysis for M2PL models using GVEM algorithms
- [D2PL\\_lrt](#) conducts DIF analysis for M2PL models using the likelihood ratio test

## Shiny app for VEMIRT

- [shinyVEMIRT](#) Run the shiny app for VEMIRT

**Author(s)**

**Maintainer:** Weicong Lyu <weiconglyu@um.edu.mo> ([ORCID](#))

Authors:

- Jiaying Xiao <jxiao6@uw.edu> ([ORCID](#))
- Ruoyi Zhu <zhux0445@uw.edu> ([ORCID](#))
- Gongjun Xu <gongjun@umich.edu> ([ORCID](#))
- Chun Wang <wang4066@uw.edu> ([ORCID](#))

**See Also**

Useful links:

- <https://MAP-LAB-UW.github.io/VEMIRT>
- <https://github.com/MAP-LAB-UW/VEMIRT>

---

C2PL\_bs

*Bootstrap Version of GVEM Confirmatory Analysis for M2PL*


---

**Description**

A bootstrap version of GVEM (i.e., GVEM-BS) can be implemented to correct the bias on item parameters and compute standard errors under confirmatory M2PL models

**Usage**

```
C2PL_bs(gvem_result, boots = 5)
```

**Arguments**

gvem_result	a list that includes exploratory or confirmatory GVEM results for M2PL models.
boots	the number of bootstrap samples; default is 5

**Value**

a list containing the following objects:

boots_a	item discrimination parameters corrected by bootstrap sampling, a $J \times K$ matrix
boots_b	item difficulty parameters corrected by bootstrap sampling, a vector of length $J$
sd_a	standard errors of item discrimination parameters, a $J \times K$ matrix
sd_b	standard errors of item difficulty parameters, a vector of length $J$

**Author(s)**

Jiaying Xiao <jxiao6@uw.edu>

**See Also**

[C2PL\\_gvem](#), [C2PL\\_iw](#)

**Examples**

```
## Not run:
gvem_result <- with(C2PL_data, C2PL_gvem(data, model))
C2PL_bs(gvem_result, boots=10)
## End(Not run)
```

C2PL\_data

*Simulated Data Set for Confirmatory M2PL Analysis***Description**

Responses are simulated based on an M2PL model with 2 factors. The true factor correlations are set as 0.8.

**Usage**

C2PL\_data

**Format**

A list of components of the data set:

data	Item responses
model	Loading indicators
params	True parameters used for generating the item responses

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

C2PL\_gvem

*Confirmatory M2PL Analysis***Description**

Confirmatory M2PL Analysis

**Usage**

```
C2PL_gvem(u, indic, max.iter = 5000, SE.est = FALSE)
```

**Arguments**

u	an $N \times J$ matrix or a data.frame that consists of binary responses of $N$ individuals to $J$ items. The missing values are coded as NA
indic	a $J \times K$ matrix or a data.frame that describes the factor loading structure of $J$ items to $K$ factors. It consists of binary values where 0 refers to the item is irrelevant with this factor, 1 otherwise
max.iter	the maximum number of iterations for the EM cycle; default is 5000
SE.est	whether to estimate SE for item parameters using the updated supplemented expectation maximization (USEM); default is FALSE

**Value**

a list containing the following objects:

ra	item discrimination parameters, a $J \times K$ matrix
rb	item difficulty parameters, vector of length $J$
reta	variational parameters $\eta(\xi)$ , a $N \times J$ matrix
reps	variational parameters $\xi$ , a $N \times J$ matrix
rsigma	population variance-covariance matrix, a $K \times K$ matrix
mu_i	mean parameter for each person, a $K \times N$ matrix
sig_i	covariance matrix for each person, a $K \times K \times N$ array
n	the number of iterations for the EM cycle
Q_mat	factor loading structure, a $J \times K$ matrix
GIC	model fit index
AIC	model fit index
BIC	model fit index
SE	Standard errors of item parameters, a $J \times (K + 1)$ matrix where the last column includes SE estimates for item difficulty parameters

**Author(s)**

Jiaying Xiao <jxiao6@uw.edu>

**See Also**

[C3PL\\_sgvm](#), [C2PL\\_bs](#), [C2PL\\_iw](#)

**Examples**

```
## Not run:
with(C2PL_data, C2PL_gvem(data, model))
## End(Not run)
```

---

C2PL\_iw

---

*Importance Weighted Version of GVEM Analysis for M2PL Models*


---

**Description**

An importance weighted version of GVEM (i.e., IW-GVEM) can be implemented to correct the bias on item parameters under M2PL models

**Usage**

```
C2PL_iw(u, gvem_result, S = 10, M = 10, max.iter = 10)
```

```
E2PL_iw(u, gvem_result, S = 10, M = 10, max.iter = 10)
```

**Arguments**

<code>u</code>	a $N \times J$ matrix or a <code>data.frame</code> that consists of binary responses of $N$ individuals to $J$ items. The missing values are coded as NA
<code>gvem_result</code>	a list that includes exploratory or confirmatory GVEM results for M2PL models.
<code>S</code>	the number of times to draw samples; default is 10
<code>M</code>	the number of samples drawn from the variational distributions; default is 10
<code>max.iter</code>	the maximum number of iterations for the EM cycle; default is 10

**Value**

a list containing the following objects:

<code>ra</code>	item discrimination parameters estimated by GVEM, a $J \times K$ matrix
<code>rb</code>	item difficulty parameters estimated by GVEM, vector of length $J$
<code>reta</code>	variational parameters $\eta(\xi)$ , a $N \times J$ matrix
<code>reps</code>	variational parameters $\xi$ , a $N \times J$ matrix
<code>rsigma</code>	population variance-covariance matrix estimated by GVEM, a $K \times K$ matrix
<code>mu_i</code>	mean parameter for each person, a $K \times N$ matrix
<code>sig_i</code>	covariance matrix for each person, a $K \times K \times N$ array
<code>n</code>	the number of iterations for the EM cycle
<code>rk</code>	factor loadings, a $J \times K$ matrix, for exploratory analysis only
<code>Q_mat</code>	factor loading structure, a $J \times K$ matrix
<code>GIC</code>	model fit index
<code>AIC</code>	model fit index
<code>BIC</code>	model fit index
<code>SE</code>	Standard errors of item parameters, a $J \times (K + 1)$ matrix where the last column includes SE estimates for item difficulty parameters, for confirmatory analysis only
<code>ur_a</code>	item discrimination parameters before conducting the rotation, a $J \times K$ matrix, for exploratory analysis only
<code>new_a</code>	item discrimination parameters estimated by IW-GVEM, a $J \times K$ matrix
<code>new_b</code>	item difficulty parameters estimated by IW-GVEM, vector of length $J$
<code>new_Sigma_theta</code>	population variance-covariance matrix estimated by IW-GVEM, a $K \times K$ matrix
<code>best_lr</code>	The learning rate used for importance sampling
<code>best_lb</code>	The lower bound value for importance sampling

**Author(s)**

Jiaying Xiao <jxiao6@uw.edu>

**See Also**

[C2PL\\_gvem](#), [E2PL\\_gvem\\_rot](#), [C2PL\\_bs](#)

## Examples

```
## Not run:
CFA_result <- with(C2PL_data, C2PL_gvem(data, model))
C2PL_iw(C2PL_data$data, CFA_result)
## End(Not run)

## Not run:
EFA_result <- with(E2PL_data_C1, E2PL_gvem_lasso(data, model, constrain = constrain, non_pen = non_pen))
E2PL_iw(E2PL_data_C1$data, EFA_result)
## End(Not run)
```

C2PL\_iw2

*IW-GVEM Algorithm for Confirmatory M2PL Analysis*

## Description

IW-GVEM Algorithm for Confirmatory M2PL Analysis

## Usage

```
C2PL_iw2(
  data,
  model = matrix(1, ncol(data)),
  criterion = "BIC",
  iter = 200,
  eps = 0.001,
  c = 1,
  S = 10,
  M = 10,
  lr = 0.1,
  SE.level = NULL
)
```

## Arguments

data	An $N \times J$ binary matrix of item responses (missing responses should be coded as NA)
model	A $J \times K$ binary matrix of loading indicators (all items load on the only dimension by default)
criterion	Information criterion for model selection, one of 'GIC' (recommended), 'BIC', or 'AIC'
iter	Maximum number of iterations
eps	Termination criterion on numerical accuracy
c	Constant for computing GIC
S	Sample size for approximating the expected lower bound
M	Sample size for approximating a tighter lower bound
lr	Learning rate for the Adam optimizer
SE.level	Accuracy level of Gaussian quadrature for mvQuad to compute standard errors (SEs are not computed if SE.level is NULL)



**Value**

An object of class `vemirt_DIF`, which is a list containing the following elements:

<code>N</code>	Number of respondents
<code>niter0</code>	Number(s) of iterations for initialization
<code>fit</code>	The only element of <code>all</code>
<code>best</code>	Equal to 1
<code>all</code>	A list of model which has one element:
<code>...\$niter</code>	Number(s) of iterations
<code>...\$SIGMA</code>	Person-level posterior covariance matrices
<code>...\$MU</code>	Person-level posterior mean vectors
<code>...\$Sigma</code>	Population covariance matrix
<code>...\$Mu</code>	Population mean vector
<code>...\$a</code>	Slopes
<code>...\$b</code>	Intercepts
<code>...\$SE.a</code>	Standard errors of <code>a</code>
<code>...\$SE.b</code>	Standard errors of <code>b</code>
<code>...\$ll</code>	Estimated lower bound of log-likelihood
<code>...\$l0</code>	Number of nonzero elements in model
<code>...\$AIC</code>	Akaike Information Criterion: $-2*ll+10*2$
<code>...\$BIC</code>	Bayesian Information Criterion: $-2*ll+10*log(N)$
<code>...\$GIC</code>	Generalized Information Criterion: $-2*ll+c*10*log(N)*log(log(N))$

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

**See Also**

[C2PL\\_gvem](#), [C2PL\\_iw](#), [D2PL\\_gvem](#), [coef.vemirt\\_DIF](#), [print.vemirt\\_DIF](#), [summary.vemirt\\_DIF](#)

**Examples**

```
## Not run:
with(C2PL_data, C2PL_iw2(data, model, SE = TRUE))
## End(Not run)
```

---

C3PL\_data

*Simulated Data Set for Confirmatory M3PL Analysis*


---

### Description

Responses are simulated based on an M3PL model with 2 factors. The true factor correlations are set as 0.8.

### Usage

C3PL\_data

### Format

A list of components of the data set:

data	Item responses
model	Loading indicators
params	True parameters used for generating the item responses

### Author(s)

Weicong Lyu <weiconglyu@um.edu.mo>

---

C3PL\_sgvem

*Stochastic GVEM for Confirmatory M3PL Analysis*


---

### Description

Stochastic GVEM for Confirmatory M3PL Analysis

### Usage

```
C3PL_sgvem(
  u,
  indic,
  samp = 50,
  forgetrate = 0.51,
  mu_b,
  sigma2_b,
  Alpha,
  Beta,
  max.iter = 5000
)
```

**Arguments**

<code>u</code>	an $N \times J$ matrix or a data.frame that consists of binary responses of $N$ individuals to $J$ items. The missing values are coded as NA
<code>indic</code>	a $J \times K$ matrix or a data.frame that describes the factor loading structure of $J$ items to $K$ factors. It consists of binary values where 0 refers to the item is irrelevant with this factor, 1 otherwise
<code>samp</code>	a subsample for each iteration; default is 50
<code>forgetrate</code>	the forget rate for the stochastic algorithm. The value should be within the range from 0.5 to 1. Default is 0.51
<code>mu_b</code>	the mean parameter for the prior distribution of item difficulty parameters
<code>sigma2_b</code>	the variance parameter for the prior distribution of item difficulty parameters
<code>Alpha</code>	the $\alpha$ parameter for the prior distribution of guessing parameters
<code>Beta</code>	the $\beta$ parameter for the prior distribution of guessing parameters
<code>max.iter</code>	the maximum number of iterations for the EM cycle; default is 5000

**Value**

a list containing the following objects:

<code>ra</code>	item discrimination parameters, a $J \times K$ matrix
<code>rb</code>	item difficulty parameters, vector of length $J$
<code>rc</code>	item guessing parameters, vector of length $J$
<code>rs</code>	variational parameters $s$ , a $N \times J$ matrix
<code>reta</code>	variational parameters $\eta(\xi)$ , a $N \times J$ matrix
<code>reps</code>	variational parameters $\xi$ , a $N \times J$ matrix
<code>rsigma</code>	population variance-covariance matrix, a $K \times K$ matrix
<code>mu_i</code>	mean parameter for each person, a $K \times N$ matrix
<code>sig_i</code>	covariance matrix for each person, a $K \times K \times N$ array
<code>n</code>	the number of iterations for the EM cycle
<code>Q_mat</code>	factor loading structure, a $J \times K$ matrix
<code>GIC</code>	model fit index
<code>AIC</code>	model fit index
<code>BIC</code>	model fit index

**Author(s)**

Jiaying Xiao <jxiao6@uw.edu>

**References**

- Cho, A. E., Wang, C., Zhang, X., & Xu, G. (2021). Gaussian variational estimation for multidimensional item response theory. *British Journal of Mathematical and Statistical Psychology*, 74, 52-85.
- Cho, A. E., Xiao, J., Wang, C., & Xu, G. (2022). Regularized Variational Estimation for Exploratory Item Factor Analysis. *Psychometrika*. <https://doi.org/10.1007/s11336-022-09874-6>

**See Also**[C2PL\\_gvem](#)**Examples**

```
## Not run:
with(C3PL_data, C3PL_sgvem(data, model, samp=50, forgetrate=0.51, mu_b=0, sigma2_b=4, Alpha=10, Beta=40))
## End(Not run)
```

---

coef.vemirt_DIF	<i>Extract Parameter Estimates from DIF 2PL Analysis</i>
-----------------	--

---

**Description**

Extract Parameter Estimates from DIF 2PL Analysis

**Usage**

```
coef(object, criterion = NULL)
```

**Arguments**

object	An object of class vemirt_DIF
criterion	Information criterion for model selection, one of 'AIC', 'BIC', 'GIC', or the constant for computing GIC, otherwise use the criterion specified when fitting the model(s)

**Author(s)**

Weicong Lyu <wlyu4@uw.edu>

**See Also**

[D2PL\\_em](#), [D2PL\\_pair\\_em](#), [D2PL\\_gvem](#), [print.vemirt\\_DIF](#), [summary.vemirt\\_DIF](#)

---

coef.vemirt_DIF_summary	<i>Extract DIF 2PL Items</i>
-------------------------	------------------------------

---

**Description**

Extract DIF 2PL Items

**Usage**

```
coef(object)
```

**Arguments**

object	An object of class vemirt_DIF_summary
--------	---------------------------------------

Author(s)

Weicong Lyu <wlyu4@uw.edu>

See Also

[summary.vemirt\\_DIF](#), [print.vemirt\\_DIF\\_summary](#)

---

coef.vemirt_FA	<i>Extract Parameter Estimates from Explanatory or Confirmatory Analysis</i>
----------------	--

---

Description

Extract Parameter Estimates from Explanatory or Confirmatory Analysis

Usage

coef(object)

Arguments

object            An object of class vemirt\_FA

Author(s)

Weicong Lyu <wlyu4@uw.edu>

See Also

[C2PL\\_gvem](#), [C2PL\\_bs](#), [C2PL\\_iw](#), [C3PL\\_sgvem](#), [E2PL\\_gvem\\_adaptlasso](#), [E2PL\\_gvem\\_lasso](#), [E2PL\\_gvem\\_rot](#), [E2PL\\_IS](#), [E3PL\\_sgvem\\_adaptlasso](#), [E3PL\\_sgvem\\_lasso](#), [E3PL\\_sgvem\\_rot](#), [print.vemirt\\_FA](#)

---

D2PL_data	<i>Simulated Data Set for DIF M2PL Analysis</i>
-----------	---

---

Description

Simulated Data Set for DIF M2PL Analysis

Usage

D2PL\_data

**Format**

A list of components of the data set:

data	Item responses
model	Loading indicators
group	Group indicators
j	Number of DIF items (the first j items have DIF)
params	A list of true parameters used for generating the item responses:
...\$a	Slopes
...\$b	Negated intercepts
...\$theta	Latent traits

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

---

D2PL\_em

---

*EM Algorithms for DIF Detection in M2PL Models*


---

**Description**

EM Algorithms for DIF Detection in M2PL Models

**Usage**

```
D2PL_em(
  data,
  model = matrix(1, ncol(data)),
  group = rep(1, nrow(data)),
  method = "EMM",
  Lambda0 = if (length(unique(group)) == 1) 0 else seq(0.1, 1, by = 0.1),
  level = 10,
  criterion = "BIC",
  iter = 200,
  eps = 0.001,
  c = 1,
  verbose = TRUE
)
```

**Arguments**

data	An $N \times J$ binary matrix of item responses (missing responses should be coded as NA)
model	A $J \times K$ binary matrix of loading indicators (all items load on the only dimension by default)
group	An $N$ dimensional vector of group indicators from 1 to G (all respondents are in the same group by default)
method	Estimation algorithm, one of 'EM' or 'EMM'
Lambda0	A vector of lambda0 values for $L_1$ penalty (lambda equals $\sqrt{N} * \text{lambda0}$ )
level	Accuracy level, either a number for mvQuad or a vector indicating the grid for each latent dimension
criterion	Information criterion for model selection, one of 'BIC' (recommended), 'AIC', or 'GIC'
iter	Maximum number of iterations
eps	Termination criterion on numerical accuracy
c	Constant for computing GIC
verbose	Whether to show the progress

**Value**

An object of class `vemirt_DIF`, which is a list containing the following elements:

N	Number of respondents
niter0	Number(s) of iterations for initialization
fit	The best (with lowest information criterion) model, which is an element of all
best	The index of fit in all
all	A list of models which has the same length as Lambda0:
...\$lambda0	Corresponding element in Lambda0
...\$lambda	$\sqrt{N} * \text{lambda0}$
...\$niter	Number(s) of iterations
...\$Sigma	Group-level covariance matrices
...\$Mu	Group-level mean vectors
...\$a	Slopes for group 1
...\$b	Intercepts for group 1
...\$gamma	D2PL parameters for the slopes
...\$beta	D2PL parameters for the intercepts
...\$ll	Log-likelihood
...\$l0	Number of nonzero D2PL parameters in gamma and beta
...\$AIC	Akaike Information Criterion: $-2*ll+10*2$
...\$BIC	Bayesian Information Criterion: $-2*ll+10*\log(N)$
...\$GIC	Generalized Information Criterion: $-2*ll+c*10*\log(N)*\log(\log(N))$

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

**See Also**

[D2PL\\_pair\\_em](#), [D2PL\\_gvem](#), [D2PL\\_lrt](#), [coef.vemirt\\_DIF](#), [print.vemirt\\_DIF](#), [summary.vemirt\\_DIF](#)

**Examples**

```
## Not run:
with(D2PL_data, D2PL_em(data, model, group))
## End(Not run)
```

D2PL\_gvem

*GVEM Algorithms for DIF Detection in M2PL Models***Description**

GVEM Algorithms for DIF Detection in M2PL Models

**Usage**

```
D2PL_gvem(
  data,
  model = matrix(1, ncol(data)),
  group = rep(1, nrow(data)),
  method = "IWGVEMM",
  Lambda0 = if (length(unique(group)) == 1) 0 else seq(0.1, 1, by = 0.1),
  criterion = "GIC",
  iter = 200,
  eps = 0.001,
  c = 1,
  S = 10,
  M = 10,
  lr = 0.1,
  verbose = TRUE
)
```

**Arguments**

data	An $N \times J$ binary matrix of item responses (missing responses should be coded as NA)
model	A $J \times K$ binary matrix of loading indicators (all items load on the only dimension by default)
group	An $N$ dimensional vector of group indicators from 1 to G (all respondents are in the same group by default)
method	Estimation algorithm, one of 'GVEM' or 'IWGVEMM'
Lambda0	A vector of lambda0 values for $L_1$ penalty (lambda equals $\sqrt{N} * \text{lambda0}$ )
criterion	Information criterion for model selection, one of 'GIC' (recommended), 'BIC', or 'AIC'
iter	Maximum number of iterations
eps	Termination criterion on numerical accuracy
c	Constant for computing GIC



S	Sample size for approximating the expected lower bound ('IWGVEMM' only)
M	Sample size for approximating a tighter lower bound ('IWGVEMM' only)
lr	Learning rate for the Adam optimizer ('IWGVEMM' only)
verbose	Whether to show the progress

**Value**

An object of class `vemirt_DIF`, which is a list containing the following elements:

N	Number of respondents
niter0	Number(s) of iterations for initialization
fit	The best (with lowest information criterion) model, which is an element of <code>all</code>
best	The index of <code>fit</code> in <code>all</code>
all	A list of models which has the same length as <code>Lambda0</code> :
...\$lambda0	Corresponding element in <code>Lambda0</code>
...\$lambda	$\sqrt{N} * \text{lambda0}$
...\$niter	Number(s) of iterations
...\$SIGMA	Person-level posterior covariance matrices
...\$MU	Person-level posterior mean vectors
...\$Sigma	Group-level covariance matrices
...\$Mu	Group-level mean vectors
...\$a	Slopes for group 1
...\$b	Intercepts for group 1
...\$gamma	D2PL parameters for the slopes
...\$beta	D2PL parameters for the intercepts
...\$RMSE	Root mean square error of fitted probability of each item for each group
...\$l1	Estimated lower bound of log-likelihood
...\$l0	Number of nonzero D2PL parameters in <code>gamma</code> and <code>beta</code>
...\$AIC	Akaike Information Criterion: $-2 * l1 + l0 * 2$
...\$BIC	Bayesian Information Criterion: $-2 * l1 + l0 * \log(N)$
...\$GIC	Generalized Information Criterion: $-2 * l1 + c * l0 * \log(N) * \log(\log(N))$

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

**See Also**

[D2PL\\_pair\\_em](#), [D2PL\\_em](#), [D2PL\\_lrt](#), [coef.vemirt\\_DIF](#), [print.vemirt\\_DIF](#), [summary.vemirt\\_DIF](#)

**Examples**

```
## Not run:
with(D2PL_data, D2PL_gvem(data, model, group))
## End(Not run)
```

D2PL\_lrt

*Likelihood Ratio Test for DIF Detection in M2PL Models***Description**

Likelihood Ratio Test for DIF Detection in M2PL Models

**Usage**

```
D2PL_lrt(data, model, group, unif = F)
```

**Arguments**

data	An $N \times J$ binary matrix of item responses
model	A $J \times K$ binary matrix of loading indicators
group	An $N$ dimensional vector of group indicators (integers from 1 to G)
unif	Whether to detect uniform D2PL only

**Value**

A list:

Sigma	Group-level posterior covariance matrices
Mu	Group-level posterior mean vectors
a	Slopes for group 1
b	Intercepts for group 1
gamma	D2PL parameters for the slopes
beta	D2PL parameters for the intercepts

**Author(s)**

Ruoyi Zhu &lt;zhux0445@uw.edu&gt;

**See Also**
[D2PL\\_em](#), [D2PL\\_pair\\_em](#), [D2PL\\_gvem](#)
**Examples**

```
## Not run:
with(D2PL_data, D2PL_lrt(data, model, group))
## End(Not run)
```

D2PL\_pair\_em

---

*EM Algorithm with ADMM for DIF Detection Using Group Pairwise Truncated  $L_1$  Penalty in 2PL Models*


---

### Description

EM Algorithm with ADMM for DIF Detection Using Group Pairwise Truncated  $L_1$  Penalty in 2PL Models

### Usage

```
D2PL_pair_em(
  data,
  group = rep(1, nrow(data)),
  Lambda0 = if (length(unique(group)) == 1) 0 else seq(0.5, 1.5, by = 0.1),
  Tau = if (length(unique(group)) == 1) 0 else c(Inf, seq(0.05, 0.3, by = 0.05)),
  rho0 = 0.5,
  level = 10,
  criterion = "BIC",
  iter = 200,
  eps = 0.001,
  c = 1,
  verbose = TRUE
)
```

### Arguments

data	An $N \times J$ binary matrix of item responses (missing responses should be coded as NA)
group	An $N$ dimensional vector of group indicators from 1 to G (all respondents are in the same group by default)
Lambda0	A vector of lambda0 values for truncated $L_1$ penalty (lambda equals $\sqrt{N} / G * \text{lambda0}$ )
Tau	A vector of tau values for truncated $L_1$ penalty (becomes $L_1$ penalty when tau equals Inf)
rho0	A value of rho for augmented Lagrangian in ADMM (tau equals $\sqrt{N} / G * \text{tau0}$ )
level	Accuracy level of Gaussian quadrature for mvQuad
criterion	Information criterion for model selection, one of 'BIC' (recommended), 'AIC', or 'GIC'
iter	Maximum number of iterations
eps	Termination criterion on numerical accuracy
c	Constant for computing GIC
verbose	Whether to show the progress

**Value**

An object of class `vemirt_DIF`, which is a list containing the following elements:

<code>N</code>	Number of respondents
<code>niter0</code>	Number(s) of iterations for initialization
<code>fit</code>	The best (with lowest information criterion) model, which is an element of <code>all</code>
<code>best</code>	The index of <code>fit</code> in <code>all</code>
<code>all</code>	A list of models which has the same length as <code>Lambda0</code> :
<code>...\$lambda0</code>	Corresponding element in <code>Lambda0</code>
<code>...\$lambda</code>	$\sqrt{N} / G * \lambda_{00}$
<code>...\$tau</code>	Corresponding element in <code>Tau</code>
<code>...\$rho0</code>	Same as <code>rho0</code> in input
<code>...\$rho</code>	$\sqrt{N} / G * \rho_{00}$
<code>...\$niter</code>	Number(s) of iterations
<code>...\$Sigma</code>	Group-level covariance matrices
<code>...\$Mu</code>	Group-level mean vectors
<code>...\$a</code>	Slopes
<code>...\$b</code>	Intercepts
<code>...\$d.a</code>	Group pairwise differences of slopes
<code>...\$d.b</code>	Group pairwise differences of intercepts
<code>...\$u.a</code>	Lagrangian multipliers of corresponding elements in <code>d.a</code>
<code>...\$u.b</code>	Lagrangian multipliers of corresponding elements in <code>d.b</code>
<code>...\$ll</code>	Log-likelihood
<code>...\$l0</code>	Number of nonzero D2PL parameters in <code>gamma</code> and <code>beta</code>
<code>...\$AIC</code>	Akaike Information Criterion: $-2*ll+10*2$
<code>...\$BIC</code>	Bayesian Information Criterion: $-2*ll+10*\log(N)$
<code>...\$GIC</code>	Generalized Information Criterion: $-2*ll+c*10*\log(N)*\log(\log(N))$

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

**See Also**

[D2PL\\_em](#), [D2PL\\_gvem](#), [D2PL\\_lrt](#), [coef.vemirt\\_DIF](#), [print.vemirt\\_DIF](#), [summary.vemirt\\_DIF](#)

**Examples**

```
## Not run:
with(D2PL_data, D2PL_pair_em(data, group, Tau = c(Inf, seq(0.01, 0.05, by = 0.01))))
## End(Not run)
```

---

E2PL_data_C1	<i>Simulated Data Set for Exploratory M2PL Analysis Under C1 Constraint</i>
--------------	---

---

**Description**

Responses are simulated based on an M2PL model with 3 factors. The true factor correlations are set as 0.5.

**Usage**

E2PL\_data\_C1

**Format**

A list of components of the data set:

data	Item responses
model	Loading indicators for (adaptive) lasso penalty
constrain	Constraint for model identification ('C1')
non_pen	Index of an item that is associated with all the factors (NULL under C1)
params	True parameters used for generating the item responses

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

---

E2PL_data_C2	<i>Simulated Data Set for Exploratory M2PL Analysis Under C2 Constraint</i>
--------------	---

---

**Description**

Responses are simulated based on an M2PL model with 3 factors. The true factor correlations are set as 0.5.

**Usage**

E2PL\_data\_C2

**Format**

A list of components of the data set:

data	Item responses
model	Loading indicators for (adaptive) lasso penalty
constrain	Constraint for model identification ('C2')
non_pen	Index of an item that is associated with all the factors
params	True parameters used for generating the item responses

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

---

E2PL\_gvem\_adaptlasso    *Exploratory M2PL Analysis with Adaptive Lasso Penalty*

---

**Description**

Exploratory M2PL Analysis with Adaptive Lasso Penalty

**Usage**

```
E2PL_gvem_adaptlasso(
  u,
  indic,
  max.iter = 5000,
  constrain = "C1",
  non_pen = NULL,
  gamma = 2
)
```

**Arguments**

u	an $N \times J$ matrix or a data.frame that consists of binary responses of $N$ individuals to $J$ items. The missing values are coded as NA
indic	a $J \times K$ matrix or a data.frame that describes the factor loading structure of $J$ items to $K$ factors. It consists of binary values where 0 refers to the item is irrelevant to this factor, and 1 otherwise. For exploratory factor analysis with adaptive lasso penalty, indic should include constraints on the a $K \times K$ sub-matrix to ensure identifiability. The remaining parts do not assume any pre-specified zero structure but instead, the appropriate lasso penalty would recover the true zero structure. Also see constrain
max.iter	the maximum number of iterations for the EM cycle; default is 5000

constrain	<p>the constraint setting: "C1" or "C2". To ensure identifiability, "C1" sets a <math>K \times K</math> sub-matrix of <math>\text{indic}</math> to be an identity matrix. This constraint anchor <math>K</math> factors by designating <math>K</math> items that load solely on each factor respectively. Note that the <math>K \times K</math> matrix does not have to appear at the top of the <math>\text{indic}</math> matrix. "C2" sets the <math>K \times K</math> sub-matrix to be a lower triangular matrix with the diagonal being ones. That is, there are test items associated with each factor for sure and they may be associated with other factors as well. Nonzero entries (in the lower triangular part) except for the diagonal entries of the sub-matrix are penalized during the estimation procedure. For instance, assume <math>K = 3</math>, then the "C2" constraint will imply the following submatrix: <math>C2 = \begin{bmatrix} 1 &amp; 0 &amp; 0 \\ 1 &amp; 1 &amp; 0 \\ 1 &amp; 1 &amp; 1 \end{bmatrix}</math>. As shown, item 1 is allowed to only load on the first factor, item 2 will for sure load on the second factor but it may also load on the first factor (hence a penalty is added on the <math>(2, 1)</math> element of "C2", i.e., <math>C2_{2,1}</math>). Item 3 will for sure load on the third factor but it may also load on the first two factors. However, note that for all remaining items their loading vector will all be <math>(1, 1, 1)</math> hence indistinguishable from the third anchor item. Therefore, we need to alert the algorithm that this third anchor item will for sure load on the third factor, and whether or not it loads on the first two factors depends on the regularization results. Therefore, we need to specify "non_pen=" to identify the <math>K</math>th anchor item. Although, "C2" is much weaker than "C1", it still ensures empirical identifiability. Default is "C1". During estimation, under both the "C1" and "C2" constraints, the population means and variances are constrained to be 0 and 1, respectively.</p>
non_pen	the index of an item that is associated with every factor under constraint "C2". For C1, the input can be NULL
gamma	a numerical value of adaptive lasso parameter. Zou (2006) recommended three values, 0.5, 1, and 2. The default value is 2.

## Value

a list containing the following objects:

ra	item discrimination parameters, a $J \times K$ matrix
rb	item difficulty parameters, vector of length $J$
reta	variational parameters $\eta(\xi)$ , a $N \times J$ matrix
reps	variational parameters $\xi$ , a $N \times J$ matrix
rsigma	population variance-covariance matrix, a $K \times K$ matrix
mu_i	mean parameter for each person, a $K \times N$ matrix
sig_i	covariance matrix for each person, a $K \times K \times N$ array
n	the number of iterations for the EM cycle
Q_mat	factor loading structure, a $J \times K$ matrix
GIC	model fit index
AIC	model fit index
BIC	model fit index
lbd	numerical value of lasso penalty parameter $\lambda$

## Author(s)

Jiaying Xiao <jxiao6@uw.edu>

## References

- Cho, A. E., Xiao, J., Wang, C., & Xu, G. (2022). Regularized Variational Estimation for Exploratory Item Factor Analysis. *Psychometrika*. <https://doi.org/10.1007/s11336-022-09874-6>
- Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*, 7, 1011418–1429.

## See Also

[E2PL\\_gvem\\_rot](#), [E2PL\\_gvem\\_lasso](#), [exampleIndic\\_efa2pl\\_c1](#), [exampleIndic\\_efa2pl\\_c2](#)

## Examples

```
## Not run:
with(E2PL_data_C1, E2PL_gvem_adaptlasso(data, model, constrain = constrain, non_pen = non_pen, gamma=2))
with(E2PL_data_C2, E2PL_gvem_adaptlasso(data, model, constrain = constrain, non_pen = non_pen, gamma=2))
## End(Not run)
```

---

E2PL\_gvem\_lasso

---

Exploratory M2PL Analysis with Lasso Penalty

---

## Description

Exploratory M2PL Analysis with Lasso Penalty

## Usage

```
E2PL_gvem_lasso(u, indic, max.iter = 5000, constrain = "C1", non_pen = NULL)
```

## Arguments

u	an $N \times J$ matrix or a data.frame that consists of binary responses of $N$ individuals to $J$ items. The missing values are coded as NA
indic	a $J \times K$ matrix or a data.frame that describes the factor loading structure of $J$ items to $K$ factors. It consists of binary values where 0 refers to the item is irrelevant with this factor, 1 otherwise. For exploratory factor analysis with lasso penalty, indic should be imposed certain constraints on the a $K \times K$ sub-matrix to ensure identifiability. The remaining parts do not assume any pre-specified zero structure but instead, the appropriate lasso penalty would recover the true zero structure. Also see constrain
max.iter	the maximum number of iterations for the EM cycle; default is 5000
constrain	the constraint setting: "C1" or "C2". To ensure identifiability, "C1" sets a $K \times K$ sub-matrix of indic to be an identity matrix. This constraint anchor $K$ factors by designating $K$ items that load solely on each factor respectively. Note that the $K \times K$ matrix does not have to appear at the top of the indic matrix. "C2" sets the $K \times K$ sub-matrix to be a lower triangular matrix with the diagonal being ones. That is, there are test items associated with each factor for sure and they may be associated with other factors as well. Nonzero entries (in the lower triangular part) except for the diagonal entries of the sub-matrix are penalized during the estimation procedure. For instance, assume $K = 3$ , then the "C2"



constraint will imply the following submatrix:  $C2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ . As shown,

item 1 is allowed to only load on the first factor, item 2 will for sure load on the second factor but it may also load on the first factor (hence a penalty is added on the (2, 1) element of "C2", i.e.,  $C2_{2,1}$ ). Item 3 will for sure load on the third factor but it may also load on the first two factors. However, note that for all remaining items their loading vector will all be (1, 1, 1) hence indistinguishable from the third anchor item. Therefore, we need to alert the algorithm that this third anchor item will for sure load on the third factor, and whether or not it loads on the first two factors depends on the regularization results. Therefore, we need to specify "non\_pen=" to identify the  $K$ th anchor item. Although, "C2" is much weaker than "C1", it still ensures empirical identifiability. Default is "C1". During estimation, under both the "C1" and "C2" constraints, the population means and variances are constrained to be 0 and 1, respectively.

non\_pen the index of an item that is associated with every factor under constraint "C2". For C1, the input can be NULL

### Value

a list containing the following objects:

ra	item discrimination parameters, a $J \times K$ matrix
rb	item difficulty parameters, vector of length $J$
reta	variational parameters $\eta(\xi)$ , a $N \times J$ matrix
reps	variational parameters $\xi$ , a $N \times J$ matrix
rsigma	population variance-covariance matrix, a $K \times K$ matrix
mu_i	mean parameter for each person, a $K \times N$ matrix
sig_i	covariance matrix for each person, a $K \times K \times N$ array
n	the number of iterations for the EM cycle
Q_mat	factor loading structure, a $J \times K$ matrix
GIC	model fit index
AIC	model fit index
BIC	model fit index
lbd	numerical value of lasso penalty parameter $\lambda$

### Author(s)

Jiaying Xiao <jxiao6@uw.edu>

### References

Cho, A. E., Xiao, J., Wang, C., & Xu, G. (2022). Regularized Variational Estimation for Exploratory Item Factor Analysis. *Psychometrika*. <https://doi.org/10.1007/s11336-022-09874-6>

### See Also

[E2PL\\_gvem\\_rot](#), [E2PL\\_gvem\\_adaptlasso](#), [exampleIndic\\_efa2pl\\_c1](#), [exampleIndic\\_efa2pl\\_c2](#)

**Examples**

```
## Not run:
with(E2PL_data_C1, E2PL_gvem_lasso(data, model, constrain = constrain, non_pen = non_pen))
with(E2PL_data_C2, E2PL_gvem_lasso(data, model, constrain = constrain, non_pen = non_pen))
## End(Not run)
```

E2PL\_gvem\_rot

*Exploratory M2PL Analysis with Post-hoc Rotation***Description**

Exploratory M2PL Analysis with Post-hoc Rotation

**Usage**

```
E2PL_gvem_rot(u, domain, max.iter = 5000, rot = "Promax")
```

**Arguments**

<code>u</code>	an $N \times J$ matrix or a data.frame that consists of binary responses of $N$ individuals to $J$ items. The missing values are coded as NA
<code>domain</code>	the number of factors
<code>max.iter</code>	the maximum number of iterations for the EM cycle; default is 5000
<code>rot</code>	the post-hoc rotation method: Promax or CF-Quartimax; default is "Promax", but may also be "cfQ" for conducting the CF-Quartimax rotation

**Value**

a list containing the following objects:

<code>ra</code>	item discrimination parameters, a $J \times K$ matrix
<code>rb</code>	item difficulty parameters, vector of length $J$
<code>reta</code>	variational parameters $\eta(\xi)$ , a $N \times J$ matrix
<code>reps</code>	variational parameters $\xi$ , a $N \times J$ matrix
<code>rsigma</code>	population variance-covariance matrix, a $K \times K$ matrix
<code>mu_i</code>	mean parameter for each person, a $K \times N$ matrix
<code>sig_i</code>	covariance matrix for each person, a $K \times K \times N$ array
<code>n</code>	the number of iterations for the EM cycle
<code>rk</code>	factor loadings, a $J \times K$ matrix
<code>Q_mat</code>	factor loading structure, a $J \times K$ matrix
<code>GIC</code>	model fit index
<code>AIC</code>	model fit index
<code>BIC</code>	model fit index
<code>ur_a</code>	item discrimination parameters before conducting the rotation, a $J \times K$ matrix

**Author(s)**

Jiaying Xiao &lt;jxiao6@uw.edu&gt;

**See Also**

[E2PL\\_gvem\\_lasso](#), [E2PL\\_gvem\\_adaptlasso](#)

**Examples**

```
## Not run:
E2PL_gvem_rot(E2PL_data_C1$data, domain=5,max.iter=3000)
E2PL_gvem_rot(E2PL_data_C1$data, domain=5,rot="cfQ")
## End(Not run)
```

---

E3PL_data_C1	<i>Simulated Data Set for Exploratory M3PL Analysis Under C1 Constraint</i>
--------------	---

---

**Description**

Responses are simulated based on an M3PL model with 3 factors. The true factor correlations are set as 0.5.

**Usage**

```
E3PL_data_C1
```

**Format**

A list of components of the data set:

data	Item responses
model	Loading indicators for (adaptive) lasso penalty
constrain	Constraint for model identification ('C1')
non_pen	Index of an item that is associated with all the factors (NULL under C1)
params	True parameters used for generating the item responses

**Author(s)**

Weicong Lyu <[weiconglyu@um.edu.mo](mailto:weiconglyu@um.edu.mo)>

---

E3PL_data_C2	<i>Simulated Data Set for Exploratory M3PL Analysis Under C2 Constraint</i>
--------------	---

---

**Description**

Responses are simulated based on an M3PL model with 3 factors. The true factor correlations are set as 0.5.

**Usage**

E3PL\_data\_C2

**Format**

A list of components of the data set:

data	Item responses
model	Loading indicators for (adaptive) lasso penalty
constrain	Constraint for model identification ('C2')
non_pen	Index of an item that is associated with all the factors
params	True parameters used for generating the item responses

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

---

E3PL_sgvem_adaptlasso	<i>Stochastic GVEM with Adaptive Lasso Penalty for Exploratory M3PL Analysis</i>
-----------------------	--

---

**Description**

Stochastic GVEM with Adaptive Lasso Penalty for Exploratory M3PL Analysis

**Usage**

```
E3PL_sgvem_adaptlasso(
  u,
  indic,
  samp = 50,
  forgetrate = 0.51,
  mu_b,
  sigma2_b,
  Alpha,
  Beta,
```

```

max.iter = 5000,
constrain = "C1",
non_pen = NULL,
gamma = 2
)

```

### Arguments

u	an $N \times J$ matrix or a data.frame that consists of binary responses of $N$ individuals to $J$ items. The missing values are coded as NA
indic	a $J \times K$ matrix or a data.frame that describes the factor loading structure of $J$ items to $K$ factors. It consists of binary values where 0 refers to the item is irrelevant with this factor, 1 otherwise. For exploratory factor analysis with lasso penalty, indic should be imposed certain constraints on the a $K \times K$ sub-matrix to ensure identifiability. The remaining parts do not assume any pre-specified zero structure but instead, the appropriate lasso penalty would recover the true zero structure. Also see constrain
samp	a subsample for each iteration; default is 50
forgetrate	the forget rate for the stochastic algorithm. The value should be within the range from 0.5 to 1. Default is 0.51
mu_b	the mean parameter for the normal prior distribution of item difficulty parameters
sigma2_b	the variance parameter for the normal prior distribution of item difficulty parameters
Alpha	the $\alpha$ parameter for the beta prior distribution of guessing parameters
Beta	the $\beta$ parameter for the beta prior distribution of guessing parameters
max.iter	the maximum number of iterations for the EM cycle; default is 5000
constrain	<p>the constraint setting: "C1" or "C2". To ensure identifiability, "C1" sets a <math>K \times K</math> sub-matrix of indic to be an identity matrix. This constraint anchor <math>K</math> factors by designating <math>K</math> items that load solely on each factor respectively. Note that the <math>K \times K</math> matrix does not have to appear at the top of the indic matrix. "C2" sets the <math>K \times K</math> sub-matrix to be a lower triangular matrix with the diagonal being ones. That is, there are test items associated with each factor for sure and they may be associated with other factors as well. Nonzero entries (in the lower triangular part) except for the diagonal entries of the sub-matrix are penalized during the estimation procedure. For instance, assume <math>K = 3</math>, then the "C2" constraint will imply the following submatrix: <math>C2 = \begin{bmatrix} 1 &amp; 0 &amp; 0 \\ 1 &amp; 1 &amp; 0 \\ 1 &amp; 1 &amp; 1 \end{bmatrix}</math>. As shown, item 1 is allowed to only load on the first factor, item 2 will for sure load on the second factor but it may also load on the first factor (hence a penalty is added on the (2, 1) element of "C2", i.e., <math>C2_{2,1}</math>). Item 3 will for sure load on the third factor but it may also load on the first two factors. However, note that for all remaining items their loading vector will all be (1, 1, 1) hence indistinguishable from the third anchor item. Therefore, we need to alert the algorithm that this third anchor item will for sure load on the third factor, and whether or not it loads on the first two factors depends on the regularization results. Therefore, we need to specify "non_pen=" to identify the <math>K</math>th anchor item. Although, "C2" is much weaker than "C1", it still ensures empirical identifiability. Default is "C1". During estimation, under both the "C1" and "C2" constraints, the population means and variances are constrained to be 0 and 1, respectively.</p>

non_pen	the index of an item which is associated with each factor to satisfy "C2". For C1, the input can be NULL
gamma	a numerical value of adaptive lasso parameter. Zou (2006) recommended three values, 0.5, 1, and 2. The default value is 2.

### Value

a list containing the following objects:

ra	item discrimination parameters, a $J \times K$ matrix
rb	item difficulty parameters, vector of length $J$
rc	item guessing parameters, vector of length $J$
rs	variational parameters $s$ , a $N \times J$ matrix
reta	variational parameters $\eta(\xi)$ , a $N \times J$ matrix
reps	variational parameters $\xi$ , a $N \times J$ matrix
rsigma	population variance-covariance matrix, a $K \times K$ matrix
mu_i	mean parameter for each person, a $K \times N$ matrix
sig_i	covariance matrix for each person, a $K \times K \times N$ array
n	the number of iterations for the EM cycle
Q_mat	factor loading structure, a $J \times K$ matrix
GIC	model fit index
AIC	model fit index
BIC	model fit index
lbd	numerical value of lasso penalty parameter $\lambda$

### Author(s)

Jiaying Xiao <jxiao6@uw.edu>

### References

- Cho, A. E., Xiao, J., Wang, C., & Xu, G. (2022). Regularized Variational Estimation for Exploratory Item Factor Analysis. *Psychometrika*. <https://doi.org/10.1007/s11336-022-09874-6>
- Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*, 7, 1011418–1429.

### See Also

[E3PL\\_sgvem\\_rot](#), [E3PL\\_sgvem\\_lasso](#), [exampleIndic\\_efa3pl\\_c1](#), [exampleIndic\\_efa3pl\\_c2](#)

### Examples

```
## Not run:
with(E3PL_data_C1, E3PL_sgvem_adaptlasso(data, model, samp=50, forgetrate=0.51, mu_b=0, sigma2_b=4, Alpha=10, Bet
with(E3PL_data_C2, E3PL_sgvem_adaptlasso(data, model, samp=50, forgetrate=0.51, mu_b=0, sigma2_b=4, Alpha=10, Bet
## End(Not run)
```

E3PL\_sgvem\_lasso

*Stochastic GVEM with Lasso Penalty for Exploratory M3PL Analysis***Description**

Stochastic GVEM with Lasso Penalty for Exploratory M3PL Analysis

**Usage**

```
E3PL_sgvem_lasso(
  u,
  indic,
  samp = 50,
  forgetrate = 0.51,
  mu_b,
  sigma2_b,
  Alpha,
  Beta,
  max.iter = 5000,
  constrain = "C1",
  non_pen = NULL
)
```

**Arguments**

u	an $N \times J$ matrix or a data.frame that consists of binary responses of $N$ individuals to $J$ items. The missing values are coded as NA
indic	a $J \times K$ matrix or a data.frame that describes the factor loading structure of $J$ items to $K$ factors. It consists of binary values where 0 refers to the item is irrelevant with this factor, 1 otherwise. For exploratory factor analysis with lasso penalty, indic should be imposed certain constraints on the a $K \times K$ sub-matrix to ensure identifiability. The remaining parts do not assume any pre-specified zero structure but instead, the appropriate lasso penalty would recover the true zero structure. Also see constrain
samp	a subsample for each iteration; default is 50
forgetrate	the forget rate for the stochastic algorithm. The value should be within the range from 0.5 to 1. Default is 0.51
mu_b	the mean parameter for the normal prior distribution of item difficulty parameters
sigma2_b	the variance parameter for the normal prior distribution of item difficulty parameters
Alpha	the $\alpha$ parameter for the beta prior distribution of guessing parameters
Beta	the $\beta$ parameter for the beta prior distribution of guessing parameters
max.iter	the maximum number of iterations for the EM cycle; default is 5000
constrain	the constraint setting: "C1" or "C2". To ensure identifiability, "C1" sets a $K \times K$ sub-matrix of indic to be an identity matrix. This constraint anchor $K$ factors by designating $K$ items that load solely on each factor respectively. Note that the $K \times K$ matrix does not have to appear at the top of the indic matrix. "C2"

sets the  $K \times K$  sub-matrix to be a lower triangular matrix with the diagonal being ones. That is, there are test items associated with each factor for sure and they may be associated with other factors as well. Nonzero entries (in the lower triangular part) except for the diagonal entries of the sub-matrix are penalized during the estimation procedure. For instance, assume  $K = 3$ , then the "C2"

constraint will imply the following submatrix:  $C2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ . As shown,

item 1 is allowed to only load on the first factor, item 2 will for sure load on the second factor but it may also load on the first factor (hence a penalty is added on the (2, 1) element of "C2", i.e.,  $C2_{2,1}$ ). Item 3 will for sure load on the third factor but it may also load on the first two factors. However, note that for all remaining items their loading vector will all be (1, 1, 1) hence indistinguishable from the third anchor item. Therefore, we need to alert the algorithm that this third anchor item will for sure load on the third factor, and whether or not it loads on the first two factors depends on the regularization results. Therefore, we need to specify "non\_pen=" to identify the  $K$ th anchor item. Although, "C2" is much weaker than "C1", it still ensures empirical identifiability. Default is "C1". During estimation, under both the "C1" and "C2" constraints, the population means and variances are constrained to be 0 and 1, respectively.

non\_pen      the index of an item which is associated with each factor to satisfy "C2". For C1, the input can be NULL

## Value

a list containing the following objects:

ra	item discrimination parameters, a $J \times K$ matrix
rb	item difficulty parameters, vector of length $J$
rc	item guessing parameters, vector of length $J$
rs	variational parameters $s$ , a $N \times J$ matrix
reta	variational parameters $\eta(\xi)$ , a $N \times J$ matrix
reps	variational parameters $\xi$ , a $N \times J$ matrix
rsigma	population variance-covariance matrix, a $K \times K$ matrix
mu_i	mean parameter for each person, a $K \times N$ matrix
sig_i	covariance matrix for each person, a $K \times K \times N$ array
n	the number of iterations for the EM cycle
Q_mat	factor loading structure, a $J \times K$ matrix
GIC	model fit index
AIC	model fit index
BIC	model fit index
lbd	numerical value of lasso penalty parameter $\lambda$

## Author(s)

Jiaying Xiao <jxiao6@uw.edu>



## References

Cho, A. E., Xiao, J., Wang, C., & Xu, G. (2022). Regularized Variational Estimation for Exploratory Item Factor Analysis. *Psychometrika*. <https://doi.org/10.1007/s11336-022-09874-6>

## See Also

[E3PL\\_sgvem\\_rot](#), [E3PL\\_sgvem\\_adaptlasso](#), [exampleIndic\\_efa3pl\\_c1](#), [exampleIndic\\_efa3pl\\_c2](#)

## Examples

```
## Not run:
with(E3PL_data_C1, E3PL_sgvem_lasso(data,model,samp=50,forgetrate=0.51,mu_b=0,sigma2_b=4,Alpha=10,Beta=40,m
with(E3PL_data_C2, E3PL_sgvem_lasso(data,model,samp=50,forgetrate=0.51,mu_b=0,sigma2_b=4,Alpha=10,Beta=40,m
## End(Not run)
```

---

E3PL\_sgvem\_rot

*Stochastic GVEM for Exploratory M3PL Analysis*


---

## Description

Stochastic GVEM for Exploratory M3PL Analysis

## Usage

```
E3PL_sgvem_rot(
  u,
  domain,
  samp = 50,
  forgetrate = 0.51,
  mu_b,
  sigma2_b,
  Alpha,
  Beta,
  max.iter = 5000,
  rot = "Promax"
)
```

## Arguments

<code>u</code>	an $N \times J$ matrix or a data.frame that consists of binary responses of $N$ individuals to $J$ items. The missing values are coded as NA
<code>domain</code>	the number of factors
<code>samp</code>	a subsample for each iteration; default is 50
<code>forgetrate</code>	the forget rate for the stochastic algorithm. The value should be within the range from 0.5 to 1. Default is 0.51
<code>mu_b</code>	the mean parameter for the prior distribution of item difficulty parameters
<code>sigma2_b</code>	the variance parameter for the prior distribution of item difficulty parameters
<code>Alpha</code>	the $\alpha$ parameter for the prior distribution of guessing parameters
<code>Beta</code>	the $\beta$ parameter for the prior distribution of guessing parameters

<code>max.iter</code>	the maximum number of iterations for the EM cycle; default is 5000
<code>rot</code>	the post-hoc rotation method: Promax or CF-Quartimax; default is "Promax", but may also be "cfQ" for conducting the CF-Quartimax rotation

### Value

a list containing the following objects:

<code>ra</code>	item discrimination parameters, a $J \times K$ matrix
<code>rb</code>	item difficulty parameters, vector of length $J$
<code>rc</code>	item guessing parameters, vector of length $J$
<code>rs</code>	variational parameters $s$ , a $N \times J$ matrix
<code>reta</code>	variational parameters $\eta(\xi)$ , a $N \times J$ matrix
<code>reps</code>	variational parameters $\xi$ , a $N \times J$ matrix
<code>rsigma</code>	population variance-covariance matrix, a $K \times K$ matrix
<code>mu_i</code>	mean parameter for each person, a $K \times N$ matrix
<code>sig_i</code>	covariance matrix for each person, a $K \times K \times N$ array
<code>n</code>	the number of iterations for the EM cycle
<code>Q_mat</code>	factor loading structure, a $J \times K$ matrix
<code>rk</code>	factor loadings, a $J \times K$ matrix
<code>GIC</code>	model fit index
<code>AIC</code>	model fit index
<code>BIC</code>	model fit index
<code>ur_a</code>	item discrimination parameters before conducting the rotation, a $J \times K$ matrix

### Author(s)

Jiaying Xiao <jxiao6@uw.edu>

### See Also

[E3PL\\_sgvem\\_lasso](#), [E3PL\\_sgvem\\_adaptlasso](#)

### Examples

```
## Not run:
E3PL_sgvem_rot(E3PL_data_C1$data, 3, samp=50, forgetrate=0.51,
mu_b=0, sigma2_b=4, Alpha=10, Beta=40, max.iter=5000, rot="Promax")
## End(Not run)
```

---

pa\_poly

Parallel analysis using polychoric correlation

---

**Description**

Identify the number of factors

**Usage**

```
pa_poly(data, n.iter = 10, figure = TRUE)
```

**Arguments**

data	a $N \times J$ matrix or a data.frame that consists of the responses of $N$ individuals to $J$ items without any missing values. The responses are binary or polytomous.
n.iter	Number of simulated analyses to perform
figure	By default, pa_poly draws an eigenvalue plot. If FALSE, it suppresses the graphic output

**Value**

pa\_poly returns a data.frame with the eigenvalues for the real data and the simulated data.

**Author(s)**

Jiaying Xiao <jxiao6@uw.edu>

**Examples**

```
## Not run:
pa_poly(C2PL_data$data, n.iter=20)
## End(Not run)
```

---

print.vemirt\_DIF

Print DIF 2PL Items by Group

---

**Description**

Print DIF 2PL Items by Group

**Usage**

```
print(x, criterion = NULL, max = 99999L, digits = 3, ...)
```

**Arguments**

x	An object of class vemirt_DIF
criterion	Information criterion for model selection, one of 'AIC', 'BIC', 'GIC', or the constant for computing GIC, otherwise use the criterion specified when fitting the model(s)

**Author(s)**

Weicong Lyu <wlyu4@uw.edu>

**See Also**

[D2PL\\_em](#), [D2PL\\_pair\\_em](#), [D2PL\\_gvem](#), [coef.vemirt\\_DIF](#), [summary.vemirt\\_DIF](#)

---

```
print.vemirt_DIF_summary
```

*Print Summary of DIF 2PL Items*

---

**Description**

Print Summary of DIF 2PL Items

**Usage**

```
print(x, max = 99999L, ...)
```

**Arguments**

x                      An object of class vemirt\_DIF\_summary

**Author(s)**

Weicong Lyu <wlyu4@uw.edu>

**See Also**

[summary.vemirt\\_DIF](#), [coef.vemirt\\_DIF\\_summary](#)

---

```
print.vemirt_FA
```

*Print Parameter Estimates from Explanatory or Confirmatory Analysis*

---

**Description**

Print Parameter Estimates from Explanatory or Confirmatory Analysis

**Usage**

```
print(x)
```

**Arguments**

x                      An object of class vemirt\_FA

**Author(s)**

Weicong Lyu <wlyu4@uw.edu>

**See Also**

[C2PL\\_gvem](#), [C2PL\\_bs](#), [C2PL\\_iw](#), [C3PL\\_sgvem](#), [E2PL\\_gvem\\_adaptlasso](#), [E2PL\\_gvem\\_lasso](#), [E2PL\\_gvem\\_rot](#), [E2PL\\_IS](#), [E3PL\\_sgvem\\_adaptlasso](#), [E3PL\\_sgvem\\_lasso](#), [E3PL\\_sgvem\\_rot](#), [coef.vemirt\\_FA](#)

shinyVEMIRT

*Shiny App for VEMIRT***Description**

Shiny App for VEMIRT

**Usage**

```
shinyVEMIRT()
```

**Author(s)**

Weicong Lyu <weiconglyu@um.edu.mo>

summary.vemirt\_DIF

*Summarize DIF 2PL Items***Description**

Summarize DIF 2PL Items

**Usage**

```
summary(x, criterion = NULL)
```

**Arguments**

criterion	Information criterion for model selection, one of 'AIC', 'BIC', 'GIC', or the constant for computing GIC, otherwise use the criterion specified when fitting the model(s)
x	An object of class vemirt_DIF

**Author(s)**

Weicong Lyu <wlyu4@uw.edu>

**See Also**

[D2PL\\_em](#), [D2PL\\_pair\\_em](#), [D2PL\\_gvem](#), [coef.vemirt\\_DIF](#), [print.vemirt\\_DIF](#), [coef.vemirt\\_DIF\\_summary](#), [print.vemirt\\_DIF\\_summary](#)

# Index

## \* datasets

- C2PL\_data, [5](#)
- C3PL\_data, [10](#)
- D2PL\_data, [13](#)
- E2PL\_data\_C1, [21](#)
- E2PL\_data\_C2, [21](#)
- E3PL\_data\_C1, [27](#)
- E3PL\_data\_C2, [28](#)

C2PL\_bs, [3](#), [4](#), [6](#), [7](#), [13](#), [37](#)

C2PL\_data, [5](#)

C2PL\_gvem, [3](#), [4](#), [5](#), [7](#), [9](#), [12](#), [13](#), [37](#)

C2PL\_iw, [3](#), [4](#), [6](#), [6](#), [9](#), [13](#), [37](#)

C2PL\_iw2, [3](#), [8](#)

C3PL\_data, [10](#)

C3PL\_sgvem, [3](#), [6](#), [10](#), [13](#), [37](#)

coef.vemirt\_DIF, [9](#), [12](#), [16](#), [17](#), [20](#), [36](#), [37](#)

coef.vemirt\_DIF\_summary, [12](#), [36](#), [37](#)

coef.vemirt\_FA, [13](#), [37](#)

D2PL\_data, [13](#)

D2PL\_em, [3](#), [12](#), [14](#), [17](#), [18](#), [20](#), [36](#), [37](#)

D2PL\_gvem, [3](#), [9](#), [12](#), [16](#), [16](#), [18](#), [20](#), [36](#), [37](#)

D2PL\_lrt, [3](#), [16](#), [17](#), [18](#), [20](#)

D2PL\_pair\_em, [3](#), [12](#), [16–18](#), [19](#), [36](#), [37](#)

E2PL\_data\_C1, [21](#)

E2PL\_data\_C2, [21](#)

E2PL\_gvem\_adaptlasso, [3](#), [13](#), [22](#), [25](#), [27](#), [37](#)

E2PL\_gvem\_lasso, [3](#), [13](#), [24](#), [24](#), [27](#), [37](#)

E2PL\_gvem\_rot, [3](#), [7](#), [13](#), [24](#), [25](#), [26](#), [37](#)

E2PL\_IS, [13](#), [37](#)

E2PL\_iw, [3](#)

E2PL\_iw (C2PL\_iw), [6](#)

E3PL\_data\_C1, [27](#)

E3PL\_data\_C2, [28](#)

E3PL\_sgvem\_adaptlasso, [3](#), [13](#), [28](#), [33](#), [34](#), [37](#)

E3PL\_sgvem\_lasso, [3](#), [13](#), [30](#), [31](#), [34](#), [37](#)

E3PL\_sgvem\_rot, [3](#), [13](#), [30](#), [33](#), [33](#), [37](#)

exampleIndic\_efa2pl\_c1, [24](#), [25](#)

exampleIndic\_efa2pl\_c2, [24](#), [25](#)

exampleIndic\_efa3pl\_c1, [30](#), [33](#)

exampleIndic\_efa3pl\_c2, [30](#), [33](#)

pa\_poly, [3](#), [35](#)

print.vemirt\_DIF, [9](#), [12](#), [16](#), [17](#), [20](#), [35](#), [37](#)

print.vemirt\_DIF\_summary, [13](#), [36](#), [37](#)

print.vemirt\_FA, [13](#), [36](#)

shinyVEMIRT, [3](#), [37](#)

summary.vemirt\_DIF, [9](#), [12](#), [13](#), [16](#), [17](#), [20](#), [36](#), [37](#)

VEMIRT (VEMIRT-package), [3](#)

VEMIRT-package, [3](#)