Package 'VEMIRT'

August 11, 2024

Type Package

Title Variational Expectation Maximization for High-Dimensional IRT Models

Version 2.2

Date 2024-08-11

Maintainer Weicong Lyu <wlyu4@uw.edu>

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

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/

Suggests knitr,
rmarkdown

VignetteBuilder knitr

Contents

`VEMIRT-package` *VEMIRT: A package for high-dimensional IRT models*

Description

VEMIRT is created to assist researchers to conduct exploratory and confirmatory multidimensional item response theory (MIRT) analysis and cooresponding 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_sgvem_rot conducts stochastic GVEM to futher imporve the computational effficiency for exploratory M3PL analysis
- E3PL_sgvem_lasso conducts M3PL Analysis with Lasso penalty
- E3PL_sgvem_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
- C3PL_sgvem conducts stochastic GVEM for confirmatory M3PL analysis

Differential item functioning analysis

- D2PL_em conducts DIF analysis for M2PL models using EM algorithms
- D2PL_gvem conducts DIF analysis for M2PL models using GVEM algorithms
- D2PL_lrt conducts DIF analysis for M2PL models using the likelihood ratio test

Author(s)

Maintainer: Weicong Lyu <wlyu4@uw.edu> (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/

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	stardard errors of item discrimination parameters, a $J \times K$ matrix
sd_b	stardard 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 a between-item M2PL model with 5 factors. The true factor correlations are set as 0.1.

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

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 ξ, 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_sgvem, 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

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

Value

a list containing the following objects:

ra	item discrimination parameters estimated by GVEM, a $J \times K$ `matrix`
rb	item difficulty parameters estimated by GVEM, vector of length J
reta	variational parameters $\eta(\xi)$, a $N \times J$ matrix
reps	variational parameters ξ, a $N \times J$ matrix
rsigma	population variance-covariance matrix estimated by GVEM, 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
rk	factor loadings, a $J \times K$ `matrix`, for exploratory analysis only
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, for confirmatory analysis only
ur_a	item discrimination parameters before conducting the rotation, a $J \times K$ `matrix`, for exploratory analysis only
new_a	item discrimination parameters estimated by IW-GVEM, a $J \times K$ `matrix`
new_b	item difficulty parameters estimated by IW-GVEM, vector of length J
new_Sigma_theta	population variance-covariance matrix estimated by IV-GVEM, a $K \times K$ matrix
best_lr	The learning rate used for importance sampling
best_lb	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)
```

C3PL_data *Simulated Data Set for Confirmatory M3PL Analysis*

Description

Responses are simulated based on a within-item M3PL model with 3 factors. The true factor correlations are set as 0.1.

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

---

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

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
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 prior distribution of item difficulty parameters
sigma2_b	the variance parameter for the prior distribution of item difficulty parameters
Alpha	the α parameter for the prior distribution of guessing parameters
Beta	the β parameter for the prior distribution of guessing parameters
max.iter	the maximum number of iterations for the EM cycle; default is 5000

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 ξ, 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

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 *Extract Parameter Estimates from DIF 2PL Analysis*

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

em_D2PL, gvemm_D2PL, lrt_D2PL, print.vemirt_DIF, summary.vemirt_DIF

---

coef.vemirt_FA *Extract Parameter Estimates from Explanatory or Confirmatory Analysis*

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` *Simulated Data Set for DIF M2PL Analysis*

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

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(0.2, 0.8, by = 0.1),
  level = 10,
  criterion = "BIC",
  iter = 200,
  eps = 0.001,
  c = 1
)
```

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) * 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

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) * lambda0
...$niter	Number(s) of iterations
...$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
...$beta	D2PL parameters for the intercepts
...$ll	Log-likelihood
...$l0	Number of nonzero D2PL parameters in gamma and beta
...$AIC	Akaike Information Criterion: -2*ll+l0*2
...$BIC	Bayesian Information Criterion: -2*ll+l0*log(N)
...$GIC	Generalized Information Criterion: -2*ll+c*l0*log(N)*log(log(N))

Author(s)

Weicong Lyu <wlyu4@uw.edu>

See Also

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(0.2, 0.8, by = 0.1),
  criterion = "GIC",
  iter = 200,
  eps = 0.001,
```

```
  c = 1,
  S = 10,
  M = 10,
  lr = 0.1
)
```

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) * 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)

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) * lambda0
...$niter	Number(s) of iterations
...$SIGMA	Person-level posterior covariance matrices
...$MU	Person-level posterior mean vectors
...$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

...$RMSE	Root mean square error of fitted probability of each item for each group
...$ll	Estimated lower bound of log-likelihood
...$l0	Number of nonzero D2PL parameters in gamma and beta
...$AIC	Akaike Information Criterion: -2*ll+l0*2
...$BIC	Bayesian Information Criterion: -2*ll+l0*log(N)
...$GIC	Generalized Information Criterion: -2*ll+c*l0*log(N)*log(log(N))

Author(s)

Weicong Lyu <wlyu4@uw.edu>

See Also

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 <zhux0445@uw.edu>

See Also

D2PL_em, D2PL_gvem

Examples

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

E2PL_data_C1 — *Simulated Data Set for Exploratory M2PL Analysis Under C1 Constraints*

Description

Responses are simulated based on a between-item M2PL model with 5 factors. The true factor correlations are set as 0.1.

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)

E2PL_data_C2 — *Simulated Data Set for Exploratory M2PL Analysis Under C2 Constraints*

Description

Responses are simulated based on a between-item M2PL model with 5 factors. The true factor correlations are set as 0.1.

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

`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`	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 Kth 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

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 ξ, 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 λ

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 identifiablity, "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 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 Kth 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 ξ, 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 λ

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

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
domain	the number of factors
max.iter	the maximum number of iterations for the EM cycle; default is 5000
rot	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:

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 ξ, 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
rk	factor loadings, a $J \times K$ matrix
Q_mat	factor loading structure, a $J \times K$ matrix
GIC	model fit index
AIC	model fit index
BIC	model fit index
ur_a	item discrimination parameters before conducting the rotation, a $J \times K$ matrix

Author(s)

Jiaying Xiao <jxiao6@uw.edu>

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 — *Simulated Data Set for Exploratory M3PL Analysis Under C1 Constraints*

Description

Responses are simulated based on a within-item M3PL model with 3 factors. The true factor correlations are set as 0.1.

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)

E3PL_data_C2 — *Simulated Data Set for Exploratory M3PL Analysis Under C2 Constraints*

Description

Responses are simulated based on a within-item M3PL model with 3 factors. The true factor correlations are set as 0.1.

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

`E3PL_sgvem_adaptlasso` *Stochastic GVEM with Adaptive Lasso Penalty for Exploratory M3PL Analysis*

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 α parameter for the beta prior distribution of guessing parameters
Beta	the β 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 identifiablity, "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 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 Kth 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
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 ξ, 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 λ

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 α parameter for the beta prior distribution of guessing parameters
Beta	the β 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 identifiablity, "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 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 Kth 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 ξ, 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 λ

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

`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`
`domain`	the number of factors
`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 prior distribution of item difficulty parameters
`sigma2_b`	the variance parameter for the prior distribution of item difficulty parameters
`Alpha`	the α parameter for the prior distribution of guessing parameters
`Beta`	the β parameter for the prior distribution of guessing parameters
`max.iter`	the maximum number of iterations for the EM cycle; default is 5000
`rot`	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:

`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 ξ, 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
`rk`	factor loadings, a $J \times K$ `matrix`

`GIC`	model fit index
`AIC`	model fit index
`BIC`	model fit index
`ur_a`	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_gvem, D2PL_lrt, coef.vemirt_DIF, summary.vemirt_DIF

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 `summary.vemirt_DIF` *Summarize DIF 2PL Items*

Description

Summarize DIF 2PL Items

Usage

```
summary(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_gvem`, `D2PL_lrt`, `coef.vemirt_DIF`, `print.vemirt_DIF`

Index