

Package ‘HighDimenCDM’

July 24, 2026

Type Package

Title Stochastic EM Algorithm for High-Dimensional Cognitive Diagnosis Models

Version 0.1.0

Author Yuxuan Mei [aut, cre],
Wenchao Ma [aut],
Kevin Wang [aut],
Gongjun Xu [aut]

Maintainer Yuxuan Mei <mei00060@umn.edu>

Description Provides stochastic expectation-maximization (stEM) algorithms for estimating high-dimensional cognitive diagnosis models. The package implements stochastic EM algorithms for cognitive diagnosis models with a large number of attributes. It includes estimation functions and example datasets for model fitting and analysis. The methods are described in Ma, W., Wang, K., and Xu, G. (Accepted). ``Parameter estimation of cognitive diagnosis models with stochastic EM algorithm." Behaviormetrika.

License GPL-3

LazyData TRUE

Depends R (>= 3.5)

Imports coda, GDINA, Rcpp (>= 0.12.1)

LinkingTo Rcpp, RcppArmadillo

Encoding UTF-8

NeedsCompilation yes

Config/roxygen2/version 8.0.0

RoxygenNote 8.0.0

Repository CRAN

Date/Publication 2026-07-24 10:10:05 UTC

Contents

batch.var 2

IPIP	2
kernel	3
simN2000K10	4
simN2000K15	5
simN2000K20	5
stEM	6
Tan2023	7
Index	9

batch.var	<i>Batch variance calculation</i>
-----------	-----------------------------------

Description

Computes variance of parameter estimates across batches in the stochastic EM algorithm.

Usage

batch.var(plist, n)

Arguments

- plist A list of parameter estimates collected across batches.
- n Number of batches

Value

A vector of batch variance estimates.

Author(s)

Wenchao Ma, The University of Minnesota, <wma@umn.edu>

IPIP	<i>International Personality Item Pool (IPIP) data</i>
------	--

Description

The IPIP-NEO-120 data are based on the public-domain personality inventory developed from the International Personality Item Pool (IPIP; Johnson, 2014).

Usage

IPIP

Format

A list of responses and Q-matrix with components:

dat Responses of 4271 participants to 120 items.

Q The 120×30 Q-matrix.

Details

The data consist of responses from 4271 participants to 120 dichotomized items and the corresponding Q-matrix relating the 120 items to 30 personality facets.

Author(s)

Yuxuan Mei, The University of Minnesota, <mei00060@umn.edu>

References

Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the ipip-neo-120. *Journal of Research in Personality*, 51, 78–89.

kernel

Gibbs Sampling Kernel for stEM Algorithm

Description

Performs Gibbs sampling updates of attribute profiles and item parameters within the stochastic EM algorithm.

Usage

```
kernel(dat, Q, J, reduced.profiles, alpha, B, ip.only = TRUE)
```

Arguments

dat	A required $N \times J$ data matrix of N examinees to J items. Missing values are currently not allowed.
Q	A required $J \times K$ item-attribute association matrix, where J is the number of items and K is the number of attributes.
J	The number of items.
reduced.profiles	A list of reduced attribute profiles for each item.
alpha	The initial latent attribute profiles.
B	The number of Gibbs sampling iterations.
ip.only	If TRUE, only item parameter estimates are returned.

Value

a list with elements

ip estimated item parameters from Gibbs sampling iterations

alpha latent attribute profiles from the final Gibbs sampling iteration

a saved latent attribute profiles across Gibbs sampling iterations

item.parm final item parameter estimates

Author(s)

Wenchao Ma, The University of Minnesota, <wma@umn.edu>

simN2000K10

Simulated CDM data with 10 attributes

Description

Simulated responses for 2000 examinees generated with the script in ‘data-raw/simN2000K10.R’ using ‘GDINA::simGDINA()’.

Usage

```
simN2000K10
```

Format

A list with components:

data A 2000×30 response matrix.

Q The 30×10 Q-matrix used for simulation.

true.params The true item category response probabilities used to generate the data.

Examples

```
data(simN2000K10)
str(simN2000K10)
```

simN2000K15

*Simulated CDM data with 15 attributes***Description**

Simulated responses for 2000 examinees generated with the script in ‘data-raw/simN2000K15.R’ using ‘GDINA::simGDINA()’.

Usage

```
simN2000K15
```

Format

A list with components:

`data` A 2000×45 response matrix.

`Q` The 45×15 Q-matrix used for simulation.

`true.params` The true item category response probabilities used to generate the data.

Examples

```
data(simN2000K15)
str(simN2000K15)
```

simN2000K20

*Simulated CDM data with 20 attributes***Description**

Simulated responses for 2000 examinees generated with the script in ‘data-raw/simN2000K20.R’ using ‘GDINA::simGDINA()’.

Usage

```
simN2000K20
```

Format

A list with components:

`data` A 2000×60 response matrix.

`Q` The 60×20 Q-matrix used for simulation.

`true.params` The true item category response probabilities used to generate the data.

Examples

```
data(simN2000K20)
str(simN2000K20)
```

stEM	<i>Stochastic EM Algorithm for High-Dimensional Cognitive Diagnosis Models</i>
------	--

Description

Estimates model parameters for cognitive diagnosis models using the stochastic expectation-maximization (stEM) algorithm with sequential Gibbs sampling.

Usage

```
stEM(
  dat,
  Q,
  item.parm = NULL,
  maxitr = 100,
  eps1 = 2,
  eps2 = 0.4,
  frac1 = 0.1,
  frac2 = 0.5,
  verbose = FALSE
)
```

Arguments

<code>dat</code>	A required $N \times J$ data matrix of N examinees to J items. Missing values are not currently supported.
<code>Q</code>	A required $J \times K$ item-attribute association matrix, where J is the number of items and K is the number of attributes.
<code>item.parm</code>	Optional initial item parameter estimates.
<code>maxitr</code>	Maximum number of iterations.
<code>eps1</code>	A positive convergence criterion used during the burn-in stage of the stochastic EM algorithm.
<code>eps2</code>	A positive convergence criterion used after the burn-in.
<code>frac1</code>	Proportion of the first part of the Markov chain used in the Geweke diagnostic.
<code>frac2</code>	Proportion of the last part of the Markov chain used in the Geweke diagnostic.
<code>verbose</code>	Logical; if TRUE, prints progress messages during the estimation process.

Value

a list with elements

catprob.parm estimated item category response probabilities

ip estimated final item parameters obtained by averaging across retained batches

alpha estimated attribute mastery profiles

alpha.list attribute mastery profile estimates from each retained batch
total.number.of.batch total number of retained batches used for estimation
final.chain final length of the Markov chain
burn.in.size the number of burn-in iterations discarded
plist estimated item parameters from each retained batch

Examples

```
dat <- simN2000K20$data
Q <- simN2000K20$Q
fit <- stEM(dat = dat, Q = Q)
fit
```

Tan2023

Mental health CDM data (Tan2023)

Description

In this study, we analyze the responses of 719 college students (34.6% men, 83.9% White, and 16.3% first-year students) to 40 items measuring mental health problems such as alcohol-related problems, anxiety, hostility, and depression.

Usage

Tan2023

Format

A list of responses and Q-matrix with components:

dat Responses of 719 participants to 40 items.

Q The 40×4 Q-matrix.

Details

The data were previously analyzed in Tan et al. (2022). The items measure four attributes: alcohol-related problems, anxiety, hostility, and depression.

Author(s)

Yuxuan Mei, The University of Minnesota, <mei00060@umn.edu>

References

Tan, Z., de la Torre, J., Ma, W., Huh, D., Larimer, M. E., & Mun, E.-Y. (2022). A tutorial on cognitive diagnosis modeling for characterizing mental health symptom profiles using existing item responses. *Prevention Science*.

Examples

```
data(Tan2023)
str(Tan2023)
fit <- stEM(
  dat = Tan2023$dat,
  Q = Tan2023$Q
)
```


Index

* datasets

IPIP, [2](#)

simN2000K10, [4](#)

simN2000K15, [5](#)

simN2000K20, [5](#)

Tan2023, [7](#)

batch.var, [2](#)

IPIP, [2](#)

kernel, [3](#)

simN2000K10, [4](#)

simN2000K15, [5](#)

simN2000K20, [5](#)

stEM, [6](#)

Tan2023, [7](#)