Package ‘variationalDCM’

March 25, 2024

Type Package

Title Variational Bayesian Estimation for Diagnostic Classification Models

Version 2.0.1

Description Enables computationally efficient parameters-estimation by variational Bayesian methods for various diagnostic classification models (DCMs). DCMs are a class of discrete latent variable models for classifying respondents into latent classes that typically represent distinct combinations of skills they possess. Recently, to meet the growing need of large-scale diagnostic measurement in the field of educational, psychological, and psychiatric measurements, variational Bayesian inference has been developed as a computationally efficient alternative to the Markov chain Monte Carlo methods, e.g., Yamaguchi and Okada (2020a) <doi:10.1007/s11336-020-09739-w>, Yamaguchi and Okada (2020b) <doi:10.3102/1076998620911934>, Yamaguchi (2020) <doi:10.1007/s41237-020-00104-w>, Oka and Okada (2023) <doi:10.1007/s11336-022-09884-4>, and Yamaguchi and Martinez (2023) <doi:10.1111/bmsp.12308>. To facilitate their applications, 'variationalDCM' is developed to provide a collection of recently-proposed variational Bayesian estimation methods for various DCMs.

Maintainer Keiichiro Hijikata <k.hijikata.1120@outlook.jp>

Depends R (>= 4.2.0)

License GPL-3

Encoding UTF-8

Imports mvtnorm, stats

Suggests knitr

VignetteBuilder knitr

RoxygenNote 7.2.2

URL https://github.com/khijikata/variationalDCM

BugReports https://github.com/khijikata/variationalDCM/issues

Config/testthat/edition 3

LazyData true

Collate 'data.R' 'dina.R' 'dino.R' 'hm_dcm.R' 'mc_dina.R' 'satu_dcm.R' 'variationalDCM.R' 'summary.R'

NeedsCompilation no

Author Keiichiro Hijikata [aut, cre],
Motonori Oka [aut] (<https://orcid.org/0000-0002-9867-8922>),
Kazuhiro Yamaguchi [aut] (<https://orcid.org/0000-0001-8011-8575>),
Kensuke Okada [aut] (<https://orcid.org/0000-0003-1663-5812>)

Repository CRAN

Date/Publication 2024-03-25 14:10:02 UTC

R topics documented:

---

dina_data_gen — *Artificial data generating function for the DINA model based on the given Q-matrix*

Description

`dina_data_gen()` returns the artificially generated item response data for the DINA model

Usage

```
dina_data_gen(Q, I, attr_cor = 0.1, s = 0.2, g = 0.2, seed = 17)
```

Arguments

`Q`	the $J \times K$ binary matrix
`I`	the number of assumed respondents
`attr_cor`	the true value of the correlation among attributes (default: 0.1)
`s`	the true value of the slip parameter (default: 0.2)
`g`	the true value of the guessing parameter (default: 0.2)
`seed`	the seed value used for random number generation (default: 17)

Value

A list including:

X the generated artificial item response data

att_pat the generated true vale of the attribute mastery pattern

References

Oka, M., & Okada, K. (2023). Scalable Bayesian Approach for the Dina Q-Matrix Estimation Combining Stochastic Optimization and Variational Inference. *Psychometrika*, 88, 302–331. doi:10.1007/s11336022098844

Examples

```
# load Q-matrix
Q = sim_Q_J80K5
sim_data = dina_data_gen(Q=Q,I=200)
```

`hm_dcm_data_gen` *Artificial data generating function for the hidden-Markov DCM based on the given Q-matrix*

Description

`hm_dcm_data_gen()` returns the artificially generated item response data for the HM-DCM

Usage

```
hm_dcm_data_gen(
  I = 500,
  Q,
  min_theta = 0.2,
  max_theta = 0.8,
  att_cor = 0.1,
  seed = 17
)
```

Arguments

`I`	the number of assumed respondents
`Q`	the $J \times K$ binary matrix
`min_theta`	the minimum value of the item parameter θ_{jht}
`max_theta`	the maximum value of the item parameter θ_{jht}
`att_cor`	the true value of the correlation among attributes (default: 0.1)
`seed`	the seed value used for random number generation (default: 17)

Value

A list including:

X the generated artificial item response data

alpha_true the generated true vale of the attribute mastery pattern, matrix form

alpha_patt_true the generated true vale of the attribute mastery pattern, string form

References

Yamaguchi, K., & Martinez, A. J. (2024). Variational Bayes inference for hidden Markov diagnostic classification models. *British Journal of Mathematical and Statistical Psychology*, 77(1), 55–79. doi:10.1111/bmsp.12308

Examples

```
indT = 3
Q = sim_Q_J30K3
hm_sim_Q = lapply(1:indT,function(time_point) Q)
hm_sim_data = hm_dcm_data_gen(Q=hm_sim_Q,I=200)
```

---

mc_dina_data_gen — *Artificial data generating function for the multiple-choice DINA model based on the given Q-matrix*

Description

`mc_dina_data_gen()` returns the artificially generated item response data for the MC-DINA model

Usage

```
mc_dina_data_gen(I, Q, att_cor = 0.1, seed = 17)
```

Arguments

`I`	the number of assumed respondents
`Q`	the $J \times K$ binary matrix
`att_cor`	the true value of the correlation among attributes (default: 0.1)
`seed`	the seed value used for random number generation (default: 17)

Value

A list including:

X the generated artificial item response data

att_pat the generated true vale of the attribute mastery pattern

References

Yamaguchi, K. (2020). Variational Bayesian inference for the multiple-choice DINA model. *Behaviormetrika*, 47(1), 159-187. doi:10.1007/s4123702000104w

Examples

```
# load a simulated Q-matrix
mc_Q = mc_sim_Q
mc_sim_data = mc_dina_data_gen(Q=mc_Q,I=200)
```

---

mc_sim_Q *Artificial Q-matrix for MC-DINA model*

Description

Artificial Q-matrix for a 30-item test measuring 5 attributes.

Usage

```
mc_sim_Q
```

Format

A matrix with components

column 1 Item number

column 2 Stem

column 3 to end attributes

References

Yamaguchi, K. (2020). Variational Bayesian inference for the multiple-choice DINA model. *Behaviormetrika*, 47(1), 159-187. doi:10.1007/s4123702000104w

sim_Q_J30K3 *Artificial Q-matrix for 30 items 3 attributes*

Description

this matrix represents an artificial Q-matrix for 30 items and 3 attributes

Usage

```
sim_Q_J30K3
```

Format

An object of class `matrix` (inherits from `array`) with 30 rows and 3 columns.

Source

artificially simulated

sim_Q_J80K5 *Artificial Q-matrix for 80 items 5 attributes*

Description

Artificial Q-matrix for a 80-item test measuring 5 attributes

Usage

```
sim_Q_J80K5
```

Format

An object of class `matrix` (inherits from `array`) with 80 rows and 5 columns.

Source

artificially simulated

variationalDCM *Variational Bayesian estimation for DCMs*

Description

variationalDCM() fits DCMs by VB algorithms.

Usage

```
variationalDCM(X, Q, model, max_it = 500, epsilon = 1e-04, verbose = TRUE, ...)

## S3 method for class 'variationalDCM'
summary(object, ...)
```

Arguments

X	$N \times J$ item response data for the DINA, DINO, MC-DINA, and saturated DCM models. Alternatively, T-length list or 3-dim array whose elements are $N \times J/T$ binary item response data matrices for the HM-DCM
Q	$J \times K$ binary Q-matrix for the DINA, DINO, and saturated DCM models. For the MC-DINA model, its size should be $J \times (K + 2)$. Alternatively, T-length list or 3-dim array whose elements are $J/T \times K$ Q-matrices for the HM-DCM
model	specify one of "dina", "dino", "mc_dina", "satu_dcm", and "hm_dcm"
max_it	Maximum number of iterations (default: 500)
epsilon	convergence tolerance for iterations (default: 1e-4)
verbose	logical, controls whether to print progress (default: TRUE)
...	additional arguments such as hyperparameter values
object	the return of the variationalDCM function and the argument of our summary function

Value

variationalDCM returns an object of class variationalDCM. We provide the summary function to summarize a result and users can check the following information:

model_params estimates of posteror means and posterior standard deviations of model parameters

attr_mastery_pat MAP etimates of attribute mastery patterns

ELBO resulting value of evidence lower bound

time time spent in computation

Methods (by generic)

- summary(variationalDCM): print summary information

variationalDCM

The `variationalDCM()` function performs recently-developed variational Bayesian inference for various DCMs. The current version can support the DINA, DINO, MC-DINA, saturated DCM, HM-DCM models. We briefly introduce additional arguments that are specific to each model.

DINA model

The DINA model has two types of model parameters: slip s_j and guessing g_j for $j = 1, \cdots, J$. We name the hyperparameters for the DINA model: `delta_0` is a L-dimensional vector, which is a hyperparameter $\boldsymbol{\delta}^0$ for the Dirichlet distribution for the class mixing parameter $\boldsymbol{\pi}$ (default: NULL). When `delta_0` is specified as `NULL`, we set $\boldsymbol{\delta}^0 = \mathbf{1}_L$. `alpha_s`, `beta_s`, `alpha_g`, and `beta_g` are positive values. They are hyperparameters $\{\alpha_s, \beta_s, \alpha_g, \beta_g\}$ that determines the shape of prior beta distribution for the slip and guessing parameters (default: NULL). When they are specified as `NULL`, they are set 1.

DINO model

The DINO model has the same model parameters and hyperparameters as the DINA model. We thus refer the readers to the DINA model.

MC-DINA model

The MC-DINA model has additional arguments `delta_0` and `a_0`. `a_0` corresponds to positive hyperparamters $\mathbf{a}^0_{jc'}$ for all j and c'. `a_0` is by default set to `NULL`, and then it is specified as 1 for all elements.

Saturated DCM

The saturated DCM is a generalized model such as the G-DINA and GDM. In the saturated DCM, we have hyperparameters $\mathbf{A}^0$ and $\mathbf{B}^0$ in addition to $\boldsymbol{\delta}^0$, which can be specified as arguments `A_0` and `B_0`. They are specified by default as `NULL`, and then we set weakly informative priors.

HM-DCM

When `model` is specified as `"hm_dcm"`, users have additional arguments `nondecreasing_attribute`, `measurement_model`, `random_block_design`, `Test_versions`, `Test_order`, `random_start`, `A_0`, `B_0`, `delta_0`, and `omega_0`. Users can accommodate the nondecreasing attribute constraint, which represents the assumption that mastered attributes are not forgotten, by setting the logical valued argument `nondecreasing_attribute` as `TRUE` (default: `FALSE`). Users can also control the measurement model by specifying `measurement_model` (default: `"general"`), and the current version can deal with the HM-general DCM (`"general"`) and HM-DINA (`"dina"`) models. This function can also handle the datasets collected by a random block design by specifying the logical valued argument `random_block_design` (default: `FALSE`). When it is specified as `TRUE`, users must enter `Test_versions` and `Test_order`. `Test_versions` is an argument indicating which version of the test each respondent has been assigned to based on a random block design, while `Test_order` indicates the sequence in which items are rearranged based on the random block design. `A_0`, `B_0`, `delta_0`, and `omega_0` correspond to hyperparameters $\mathbf{A}^0$, $\mathbf{B}^0$, $\boldsymbol{\delta}^0$, and $\boldsymbol{\Omega}^0$. $\boldsymbol{\Omega}^0$ is nonnegative hyperparameters of Dirichlet distributions for attribute transition probabilities. `omega_0` is by default set to `NULL`, and then we set $\boldsymbol{\Omega}^0 = \mathbf{1}_L \mathbf{1}_L^\top$.

References

Yamaguchi, K., & Okada, K. (2020). Variational Bayes inference for the DINA model. *Journal of Educational and Behavioral Statistics*, 45(5), 569-597. doi:10.3102/1076998620911934

Yamaguchi, K. (2020). Variational Bayesian inference for the multiple-choice DINA model. *Behaviormetrika*, 47(1), 159-187. doi:10.1007/s4123702000104w

Yamaguchi, K., Okada, K. (2020). Variational Bayes Inference Algorithm for the Saturated Diagnostic Classification Model. *Psychometrika*, 85(4), 973–995. doi:10.1007/s1133602009739w

Yamaguchi, K., & Martinez, A. J. (2024). Variational Bayes inference for hidden Markov diagnostic classification models. *British Journal of Mathematical and Statistical Psychology*, 77(1), 55–79. doi:10.1111/bmsp.12308

Examples

```
# fit the DINA model
Q = sim_Q_J80K5
sim_data = dina_data_gen(Q=Q,I=200)
res = variationalDCM(X=sim_data$X, Q=Q, model="dina")
summary(res)
```

Index