Package 'REMLA'

December 5, 2024

Title Robust Expectation-Maximization Estimation for Latent Variable Models

Version 1.2.0

Date 2024-11-30

Description Traditional latent variable models assume that the population is homogeneous, meaning that all individuals in the population are assumed to have the same latent structure. However, this assumption is often violated in practice given that individuals may differ in their age, gender, socioeconomic status, and other factors that can affect their latent structure. The robust expectation maximization (REM) algorithm is a statistical method for estimating the parameters of a latent variable model in the presence of population heterogeneity as recommended by Nieser & Cochran (2023) <doi:10.1037/met0000413>. The REM algorithm is based on the expectation-maximization (EM) algorithm, but it allows for the case when all the data are generated by the assumed data generating model.

License GPL (>= 3)

URL https://github.com/knieser/REM

Depends R (>= 4.0), GPArotation,geex

Imports stats

Encoding UTF-8

RoxygenNote 7.3.2

Suggests knitr, lavaan, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Author Bryan Ortiz-Torres [aut, cre],
Kenneth Nieser [aut] (<https://orcid.org/0000-0001-6003-1296>)

Maintainer Bryan Ortiz-Torres <bortiztorres@wisc.edu>

Repository CRAN

Date/Publication 2024-12-05 20:50:02 UTC

Contents

`controlREM` *Control parameters for REM package*

Description

Control parameters for REM package

Usage

```
controlREM(
  steps = 25,
  tol = 1e-06,
  maxiter = 1000,
  min_weights = 1e-30,
  max_ueps = 0.3,
  chk_gamma = 0.9,
  n = 20000
)
```

Arguments

`steps`	number of steps in binary search for optimal epsilon value (default = 25)
`tol`	tolerance parameter to check for convergence of EM and REM algorithm (default = 1e-6)
`maxiter`	maximum number iterations of EM and REM algorithm (default = 1e3)
`min_weights`	lower bound for the individual weights estimated by REM (default = 1e-30)
`max_ueps`	percentile of the distribution of likelihood values to use as the maximum epsilon value to consider
`chk_gamma`	gamma value used when searching for epsilon
`n`	sample size of simulated data used when checking heuristic criterion in the epsilon search

Value

control parameters used in the REM package (steps, tol, maxiter, min_weights, ueps, n).

Author(s)

Bryan Ortiz-Torres (bortiztorres@wisc.edu); Kenneth Nieser (nieser@stanford.edu)

References

Nieser, K. J., & Cochran, A. L. (2023). Addressing heterogeneous populations in latent variable settings through robust estimation. Psychological methods, 28(1), 39.

See Also

`REM_EFA()`, `REM_CFA()`

`REM_CFA` *Robust Estimation Maximization Estimates for Confirmatory Factor Analysis*

Description

This function uses the robust expectation maximization (REM) algorithm to estimate the parameters of a confirmatory factor analysis model as suggested by Nieser & Cochran (2023).

Usage

```
REM_CFA(X, delta = 0.05, model = NA, ctrREM = controlREM())
```

Arguments

`X`	data to analyze; should be a data frame or matrix
`delta`	hyperparameter between 0 and 1 that captures the researcher's tolerance of incorrectly down-weighting data from the model (default = 0.05).
`model`	string variable that contains each structural equation in a new line where equalities are denoted by the symbol "~".
`ctrREM`	control parameters (default: (steps = 25, tol = 1e-6, maxiter = 1e3, min_weights = 1e-30, max_ueps = 0.3, chk_gamma = 0.9, n = 2e4))

Value

REM_CFA returns an object of class "REM". The function `summary()` is used to obtain estimated parameters from the model. An object of class "REM" in Confirmatory Factor Analysis is a list of outputs with four different components: the matched call (call), estimates using traditional expectation maximization (EM_output), estimates using robust expectation maximization (REM_output), and a summary table (summary_table). The list contains the following components:

`call`	match call
`model`	model frame
`delta`	hyperparameter between 0 and 1 that captures the researcher's tolerance of incorrectly down-weighting data from the model
`k`	number of factors

`constraints`	p x k matrix of zeros and ones denoting the factors (rows) and observed variables (columns)
`epsilon`	hyperparameter on the likelihood scale
`AIC_rem`	Akaike Information Criterion
`BIC_rem`	Bayesian Information Criterion
`mu`	item intercepts
`lambda`	factor loadings
`psi`	unique variances of items
`gamma`	average weights
`weights`	estimated REM weights
`ind_lik`	likelihood value for each individual
`lik_rem`	joint log-likelihood evaluated at REM estimates
`lik`	joint log-likelihood evaluated at EM estimates
`summary_table`	summary of EM and REM estimates, SEs, Z statistics, p-values, and 95% confidence intervals

Author(s)

Bryan Ortiz-Torres (bortiztorres@wisc.edu); Kenneth Nieser (nieser@stanford.edu)

References

Nieser, K. J., & Cochran, A. L. (2023). Addressing heterogeneous populations in latent variable settings through robust estimation. Psychological methods, 28(1), 39.

See Also

`REM_EFA()`, `summary.REMLA()`

Examples

```
# CFA of Holzinger-Swineford dataset
library(lavaan)
df <- HolzingerSwineford1939
data = df[,-c(1:6)]

model <- "Visual  =~  x1 + x2 + x3
          Textual =~  x4 + x5 + x6
          Speed   =~  x7 + x8 + x9"

model_CFA = REM_CFA(X = data, model = model)
summary(model_CFA)
```

REM_EFA *Robust Estimation Maximization for Exploratory Factor Analysis*

Description

This function uses the robust expectation maximization (REM) algorithm to estimate the parameters of an exploratory factor analysis model as suggested by Nieser & Cochran (2023).

Usage

```
REM_EFA(X, k_range, delta = 0.05, rotation = "oblimin", ctrREM = controlREM())
```

Arguments

X	data to analyze; should be a data frame or matrix
k_range	vector of the number of factors to consider
delta	hyperparameter between 0 and 1 that captures the researcher's tolerance of incorrectly down-weighting data from the model (default = 0.05)
rotation	factor rotation method (default = 'oblimin'); 'varimax' is the only other available option at this time
ctrREM	control parameters (default: (steps = 25, tol = 1e-6, maxiter = 1e3, min_weights = 1e-30, max_ueps = 0.3, chk_gamma = 0.9, n = 2e4))

Value

REM_EFA returns an object of class "REM". The function `summary()` is used to obtain estimated parameters from the model. An object of class "REM" in Exploratory Factor Analysis is a list of outputs with four different components for each number of factor: the matched call (call), estimates using traditional expectation maximization (EM_output), estimates using robust expectation maximization (REM_output), and a summary table (summary_table). The list contains the following components:

call	match call
model	model frame
k	number of factors
constraints	p x k matrix of zeros and ones denoting the factors (rows) and observed variables (columns)
epsilon	hyperparameter on the likelihood scale
AIC_rem	Akaike information criterion based on REM estimates
BIC_rem	Bayesian information criterion based on REM estimates
mu	item intercepts
lambda	factor loadings
psi	unique variances of items

phi	factor covariance matrix
gamma	average weight
weights	estimated REM weights
ind_lik	likelihood value for each individual
lik_rem	joint log-likelihood evaluated at REM estimates
lik	joint log-likelihood evaluated at EM estimates
mu.se	standard errors of items intercepts
lambda.se	standard errors of factor loadings
psi.se	standard errors of unique variances of items
gamma.se	standard error of gamma
summary_table	summary of EM and REM estimates, SEs, Z statistics, p-values, and 95% confidence intervals

The summary function can be used to obtain estimated parameters from the optimal model based on the BIC from the EM and REM algorithms.

Author(s)

Bryan Ortiz-Torres (bortiztorres@wisc.edu); Kenneth Nieser (nieser@stanford.edu)

References

Nieser, K. J., & Cochran, A. L. (2023). Addressing heterogeneous populations in latent variable settings through robust estimation. Psychological methods, 28(1), 39.

See Also

REM_CFA(), summary.REMLA() for more detailed summaries, GPArotation::oblimin() and varimax() for details on the rotation

Examples

```
# EFA of Holzinger-Swineford dataset
library(lavaan)
df <- HolzingerSwineford1939
data = df[,-c(1:6)]

model_EFA = REM_EFA(X = data, k_range = 1:3)
summary(model_EFA)
```

summary.REMLA *Summary for Robust Estimation Maximization*

Description

Summary method for class "REMLA".

Usage

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

Arguments

`object`	an object of class "REMLA", usually a result of a call to REM_EFA.
`...`	further arguments passed to or from other methods.

Value

The summary.REM function returns estimated parameters from the optimal model based on the BIC from the EM and REM algorithms.

Output include:

`optimal`	optimal number of factors based on BIC
`mu`	intercept
`lambda`	loadings
`psi`	variance
`indk_lik`	likelihood value for each individual
`epsilon`	hyperparameter on the likelihood scale
`diff`	differences between EM and REM

Author(s)

Bryan Ortiz-Torres (bortiztorres@wisc.edu); Kenneth Nieser (nieser@stanford.edu)

References

Nieser, K. J., & Cochran, A. L. (2023). Addressing heterogeneous populations in latent variable settings through robust estimation. Psychological methods, 28(1), 39.

See Also

`REM_EFA()`, `REM_CFA()`, `summary()`.

Index