Package ‘CFM’

May 7, 2026

Type Package

Title Analyzing Censored Factor Models

Version 0.8.0

Description Provides generation and estimation of censored factor models for high-dimensional data with censored errors (normal, t, logistic). Includes Sparse Orthogonal Principal Components (SOPC), and evaluation metrics. Based on Guo G. (2023) <doi:10.1007/s00180-022-01270-z>.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

Depends R (>= 3.5.0)

Imports stats, MASS, psych, matrixcalc, crch, Matrix, pracma, RSpectra, kernlab, mvtnorm, robustbase, truncnorm

Suggests testthat (>= 3.0.0), ggplot2

NeedsCompilation no

RoxygenNote 7.3.3

Author Guangbao Guo [aut, cre], Tong Meng [aut]

Maintainer Guangbao Guo <ggb11111111@163.com>

Repository CRAN

Date/Publication 2025-12-07 07:10:02 UTC

Contents

Aids2 2
bcdata 3
censored_factor_model 3
censored_factor_models 4
censored_kernel_factor_model 5
CFM 6
FanPC.CFM 7

incremental_censored_factor_model 8
PC2.CFM 9
PPC2.CFM 10
weighted_censored_factor_model 12
yoghurt 12

Index **14**

`Aids2` *Aids2 Dataset*

Description

Australian AIDS survival data containing information on patients diagnosed with AIDS in Australia.

Usage

```
data(Aids2)
```

Format

A data frame with 2843 observations and 7 variables:

state State in Australia: NSW, Other, QLD, VIC

sex Sex of the patient

diag Date of diagnosis

death Date of death

status Status at the end of the study: A (alive), D (dead)

T.categ Transmission category

age Age at diagnosis

Source

Australian National AIDS Registry

Examples

```
data(Aids2)
head(Aids2)
summary(Aids2)
```

bcdata *Boundary Condition Dataset*

Description

A dataset containing boundary condition data for computational fluid mechanics simulations.

Usage

```
data(bcdata)
```

Format

A data frame with variables:

time Time variable (seconds)

pressure Pressure values (Pa)

temperature Temperature values (K)

velocity Velocity components (m/s)

Source

Experimental measurements or numerical simulations

Examples

```
data(bcdata)
head(bcdata)
summary(bcdata)
```

censored_factor_model *Robust Censored Factor Model*

Description

Implementation of robust censored factor model using robust PCA initialization for handling left-censored data.

Usage

```
censored_factor_model(
  X,
  m,
  max_iter = 100,
  tol = 1e-04,
  nugget = 1e-06,
  alpha = 0.75
)
```

Arguments

`X`	Data matrix (n x p)
`m`	Number of factors
`max_iter`	Maximum number of ECM iterations (default: 100)
`tol`	Convergence tolerance (default: 1e-4)
`nugget`	Numerical stability term (default: 1e-6)
`alpha`	Robustness parameter for MCD (default: 0.75)

Value

A list containing model results

`censored_factor_models` *Basic censored-factor data simulator*

Description

Generates multivariate data that follow a latent factor structure with censored errors (Normal, Student-t or Logistic).

Usage

```
censored_factor_models(
  n,
  p,
  m,
  distribution = c("normal", "t", "logistic"),
  df = NULL,
  seed = NULL
)
```

Arguments

`n`	Sample size (> 0).
`p`	Number of observed variables (> 0).
`m`	Number of latent factors (< p).
`distribution`	Error distribution: "normal" (default), "t", "logistic".
`df`	Degrees of freedom when distribution = "t".
`seed`	Optional random seed.

Value

A list with components:

`data`	numeric n × p matrix of observations
`loadings`	p × m factor loadings matrix
`uniqueness`	p × p diagonal uniqueness matrix
`KMO`	KMO measure of sampling adequacy
`Bartlett_p`	p-value of Bartlett's test
`distribution`	error distribution used
`seed`	random seed

Examples

```
set.seed(2025)
obj <- censored_factor_models(200, 6, 2)
psych::KMO(obj$data)
```

---

censored_kernel_factor_model

Kernel Censored Factor Model

Description

Implementation of kernel-based censored factor model using kernel PCA initialization for nonlinear factor analysis with censored data.

Usage

```
censored_kernel_factor_model(
  X,
  m,
  kernel_type = "rbf",
  gamma = NULL,
  max_iter = 100,
  tol = 1e-04,
  nugget = 1e-06
)
```

Arguments

`X`	Data matrix (n x p)
`m`	Number of factors
`kernel_type`	Kernel type: "rbf" or "linear" (default: "rbf")
`gamma`	Gamma parameter for RBF kernel (default: 1/p)

max_iter	Maximum number of ECM iterations (default: 100)
tol	Convergence tolerance (default: 1e-4)
nugget	Numerical stability term (default: 1e-6)

Value

A list containing model results

CFM *Censored Factor Models Data Generation*

Description

Generate multivariate data that follow a latent factor structure with censoring errors drawn from Normal, Student-t or Logistic distributions. Convenience wrapper around rcnorm, rct, and rclogis.

Usage

```
CFM(n, p, m, cens.dist = c("normal", "t", "logistic"), df = 5, seed = NULL)
```

Arguments

n	sample size ($n \times 1$ observations).
p	number of manifest variables.
m	number of latent factors.
cens.dist	censoring error distribution: "normal", "t", or "logistic".
df	degrees of freedom when cens.dist = "t".
seed	optional random seed for reproducibility.

Value

A named list with components:

data	numeric $n \times p$ matrix of observations.
F	factor scores matrix ($n \times m$).
A	factor loadings matrix ($p \times m$).
D	unique variances diagonal matrix ($p \times p$).

Examples

```
set.seed(2025)
# Normal censoring
obj <- CFM(n = 200, p = 10, m = 3, cens.dist = "normal")
head(obj$data)

# t-censoring with 6 d.f.
obj <- CFM(n = 300, p = 12, m = 4, cens.dist = "t", df = 6)
psych::KMO(obj$data)
```

FanPC.CFM *Censored Factor Analysis via Principal Component (FanPC, pure R)*

Description

Censored Factor Analysis via Principal Component (FanPC, pure R)

Usage

```
FanPC.CFM(
  data,
  m,
  A = NULL,
  D = NULL,
  p = NULL,
  cens.dist = c("normal", "t", "logistic"),
  df = NULL,
  cens.method = c("winsorise", "em"),
  cens_prop = 0.01,
  surv.obj = NULL,
  ctrl = NULL,
  verbose = NULL
)
```

Arguments

data	Numeric matrix or data frame of dimension $n \times p$.
m	Number of factors (< p).
A	Optional true loading matrix, used only for error calculation.
D	Optional true unique-variance diagonal matrix, used only for error calculation.
p	Number of variables (deprecated; detected automatically).
cens.dist	Error distribution, reserved for future use.
df	Degrees of freedom, reserved for future use.

`cens.method`	Censoring handling method; currently only `"winsorise"` is implemented. Defaults to `"winsorise"`.
`cens_prop`	Winsorisation proportion, default 0.01.
`surv.obj`	Reserved for future use.
`ctrl`	Reserved for future use.
`verbose`	Reserved for future use.

Value

AF Estimated loading matrix, p × m.

DF Estimated unique-variance diagonal matrix, p × p.

MSESigmaA Mean squared error of loadings (if A is provided).

MSESigmaD Mean squared error of unique variances (if D is provided).

LSigmaA Relative error of loadings (if A is provided).

LSigmaD Relative error of unique variances (if D is provided).

Examples

```
library(CFM)
obj <- CFM(n = 500, p = 10, m = 2, cens.dist = "normal")
res <- FanPC.CFM(obj$data, m = 2, A = obj$A, D = obj$D, cens.method = "winsorise")
print(res$MSESigmaA)
```

`incremental_censored_factor_model`

Incremental Censored Factor Model

Description

Implementation of incremental censored factor model for streaming data with left-censored observations. Processes data in batches.

Usage

```
incremental_censored_factor_model(
  batch_data,
  m,
  max_iter = 100,
  tol = 1e-04,
  nugget = 1e-06
)
```

Arguments

batch_data	List of data matrices (batches)
m	Number of factors
max_iter	Maximum number of ECM iterations (default: 100)
tol	Convergence tolerance (default: 1e-4)
nugget	Numerical stability term (default: 1e-6)

Value

A list containing model results

PC2.CFM — *PC2 for censored factor models (Top-2 principal components, pure R)*

Description

PC2 for censored factor models (Top-2 principal components, pure R)

Usage

```
PC2.CFM(
  data,
  m,
  A = NULL,
  D = NULL,
  p = NULL,
  cens.dist = c("normal", "t", "logistic"),
  df = NULL,
  cens.method = c("winsorise", "em"),
  cens_prop = 0.01,
  surv.obj = NULL,
  ctrl = NULL,
  verbose = NULL
)
```

Arguments

data	Numeric matrix or data frame of dimension $n \times p$.
m	Number of factors (< p).
A	Optional true loading matrix, used only for error calculation.
D	Optional true unique-variance diagonal matrix, used only for error calculation.
p	Number of variables (deprecated; detected automatically).
cens.dist	Error distribution, reserved for future use.
df	Degrees of freedom, reserved for future use.

`cens.method`	Censoring handling method; currently only `"winsorise"` is implemented. Defaults to `"winsorise"`.
`cens_prop`	Winsorisation proportion, default 0.01.
`surv.obj`	Reserved for future use.
`ctrl`	Reserved for future use.
`verbose`	Reserved for future use.

Value

AF Estimated loading matrix, $p \times 2$.

DF Estimated unique-variance diagonal matrix, $p \times p$.

MSESigmaA Mean squared error of loadings (if A is provided).

MSESigmaD Mean squared error of unique variances (if D is provided).

LSigmaA Relative error of loadings (if A is provided).

LSigmaD Relative error of unique variances (if D is provided).

Examples

```
library(CFM)
obj <- CFM(n = 500, p = 12, m = 2, cens.dist = "normal")
res <- PC2.CFM(obj$data, A = obj$A, D = obj$D)
print(res$MSESigmaA)
```

`PPC2.CFM` *PPC2 for censored factor models (Top-2 principal components, pure R)*

Description

PPC2 for censored factor models (Top-2 principal components, pure R)

Usage

```
PPC2.CFM(
  data,
  m,
  A = NULL,
  D = NULL,
  p = NULL,
  cens.dist = c("normal", "t", "logistic"),
  df = NULL,
  cens.method = c("winsorise", "em"),
  cens_prop = 0.01,
```

```
  surv.obj = NULL,
  ctrl = NULL,
  verbose = NULL
)
```

Arguments

`data`	Numeric matrix or data frame of dimension $n \times p$.
`m`	Number of factors (< p).
`A`	Optional true loading matrix, used only for error calculation.
`D`	Optional true unique-variance diagonal matrix, used only for error calculation.
`p`	Number of variables (deprecated; detected automatically).
`cens.dist`	Error distribution, reserved for future use.
`df`	Degrees of freedom, reserved for future use.
`cens.method`	Censoring handling method; currently only `"winsorise"` is implemented. Defaults to `"winsorise"`.
`cens_prop`	Winsorisation proportion, default 0.01.
`surv.obj`	Reserved for future use.
`ctrl`	Reserved for future use.
`verbose`	Reserved for future use.

Value

AF Estimated loading matrix, p × 2.

DF Estimated unique-variance diagonal matrix, p × p.

MSESigmaA Mean squared error of loadings (if A is provided).

MSESigmaD Mean squared error of unique variances (if D is provided).

LSigmaA Relative error of loadings (if A is provided).

LSigmaD Relative error of unique variances (if D is provided).

Examples

```
library(CFM)
obj <- CFM(n = 500, p = 12, m = 2, cens.dist = "normal")
res <- PPC2.CFM(obj$data, A = obj$A, D = obj$D, cens.method = "winsorise")
print(res$MSESigmaA)
```

weighted_censored_factor_model *Weighted Censored Factor Model*

Description

Implementation of weighted censored factor model with weighted PCA initialization for handling left-censored data with observation weights.

Usage

```
weighted_censored_factor_model(
  X,
  m,
  weights = NULL,
  max_iter = 100,
  tol = 1e-04,
  nugget = 1e-06
)
```

Arguments

X	Data matrix (n x p)
m	Number of factors
weights	Observation weights vector of length n (optional)
max_iter	Maximum number of ECM iterations (default: 100)
tol	Convergence tolerance (default: 1e-4)
nugget	Numerical stability term (default: 1e-6)

Value

A list containing model results

yoghurt *Yoghurt Dataset*

Description

A dataset containing experimental or survey data related to yogurt.

Usage

```
data(yoghurt)
```

Format

A data frame with the following columns:

adult Description of adult variable

fadult Description of fadult variable

left Description of left variable

right Description of right variable

Details

This dataset contains experimental data collected from yogurt-related studies. The specific meaning of each variable should be documented based on the original study.

Source

Original data source should be specified here.

Examples

```
# Load the data
data(yoghurt)

# Examine data structure
str(yoghurt)

# View first few rows
head(yoghurt)

# Basic summary statistics
summary(yoghurt)
```

Index

∗ **datasets**
    yoghurt, 12

Aids2, 2

bcdata, 3

censored_factor_model, 3
censored_factor_models, 4
censored_kernel_factor_model, 5
CFM, 6

FanPC.CFM, 7

incremental_censored_factor_model, 8

PC2.CFM, 9
PPC2.CFM, 10

rclogis, *6*
rcnorm, *6*
rct, *6*

weighted_censored_factor_model, 12

yoghurt, 12