Package 'BFM'

May 18, 2026

Type Package

Title Beta Factor Model

Version 0.2.11

Author Guangbao Guo [aut, cre],
Jiahui Feng [aut]

Maintainer Guangbao Guo <ggb11111111@163.com>

Description
Provides tools for factor analysis in financial and econometric settings under Beta factor models. It includes functions to simulate factor-model data with Beta-distributed idiosyncratic components (e.g., standard Beta, scaled Beta, and truncated Beta distributions) and to conduct model diagnostic assessments such as likelihood ratio tests for factor number selection and goodness-of-fit tests for Beta distribution assumptions. Estimation routines encompass maximum likelihood estimation for finite-dimensional Beta factor models, regularized Beta factor analysis for high-dimensional datasets, and shrinkage-based estimation for robust Beta factor loading recovery in noisy or incomplete data environments. The package's methodological framework is detailed in Guo G. (2023) <doi:10.1007/s00180-022-01270-z>.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 3.5.0)

Suggests testthat (>= 3.0.0), spelling, betareg, zoib

NeedsCompilation no

LazyData true

RoxygenNote 7.3.3

Imports MASS, psych, stats

Language en-US

Repository CRAN

Date/Publication 2026-05-18 18:20:13 UTC

Contents

`AlcoholUse` *California Alcohol Use Data*

Description

A county-level monthly alcohol use dataset from California students (grades 7-11, 2008-2010). The response variable `Percentage` is a proportion (0 < Percentage < 1), suitable for zero-inflated beta regression.

Usage

```
AlcoholUse
```

Format

A data frame with multiple rows and variables:

Percentage numeric: percentage of students who drank alcohol

Grade factor: student grade level

Gender factor: student gender

MedDays numeric: mid-point of days bucket

Days numeric: days bucket

County factor: county identifier

A data frame with 44 rows and 4 variables:

accuracy numeric: proportion of correct responses in a reading task

accuracy1 numeric: transformed accuracy measure

dyslexia factor: dyslexia status (levels: "yes", "no")

iq numeric: IQ score

Source

http://www.kidsdata.org Reading Skills Data

A dataset from Smithson and Verkuilen (2006) on reading accuracy, dyslexia status, and IQ scores. The response variable accuracy is a proportion (0 < accuracy < 1), suitable for beta regression.

Smithson, M. & Verkuilen, J. (2006). *A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables*. https://psycnet.apa.org/doi/10.1037/1082-989X.11.1.54

Examples

```
data(AlcoholUse)
str(AlcoholUse)
```

BFM *The BFM function is to generate Beta Factor Models data.*

Description

The function supports various distribution types for generating the data.

Usage

```
BFM(n, p, m, mub, phib, distribution_type)
```

Arguments

n	Sample size.
p	Sample dimensionality.
m	Number of factors.
mub	Mean parameter for Beta distribution (numeric vector or scalar, 0 < mub < 1).
phib	Precision parameter for Beta distribution (positive numeric vector or scalar).
distribution_type	Type of Beta distribution.

Value

A list containing:

data	Generated BFM data matrix (n rows, p columns).
A	A matrix representing the factor loadings.
D	Diagonal matrix of unique variances.
kmo	Kaiser-Meyer-Olkin sampling adequacy measure.
bartlett	Bartlett's test of sphericity.

Examples

```
n <- 1000
p <- 10
m <- 5
mub <- runif(p, 0.2, 0.8)
phib <- runif(p, 5, 30)
dist_type <- "Elliptical Distribution"
X <- BFM(n, p, m, mub, phib, dist_type)
```

`calculate_errors` *Calculate Errors for Factor Analysis Estimates*

Description

This function calculates the Mean Squared Error (MSE) and relative error for factor loadings and uniqueness estimates.

Usage

```
calculate_errors(data, A, D, estimation_results)
```

Arguments

`data`	Matrix of BFM data.
`A`	Matrix of true factor loadings.
`D`	Matrix of true uniquenesses (diagonal matrix).
`estimation_results`	A list containing `A_hat` (estimated loadings) and `D_hat` (estimated uniquenesses).

Value

A named vector containing:

`MSEA`	Mean Squared Error for factor loadings.
`MSED`	Mean Squared Error for uniqueness estimates.
`LSA`	Relative error for factor loadings.
`LSD`	Relative error for uniqueness estimates.

Examples

```
set.seed(123)
n <- 10
p <- 5
A <- matrix(runif(p * p, -1, 1), nrow = p)
D <- diag(runif(p, 1, 2))
data <- matrix(runif(n * p), nrow = n)
estimation_results <- list(A_hat = A, D_hat = D)
errors <- calculate_errors(data, A, D, estimation_results)
print(errors)
```

FoodExpenditure *Household Food Expenditure Data*

Description

A dataset from Griffiths, Hill, and Judge (1993) on household food expenditure, income, and household size. The response variable `food` is a proportion ($0 < \text{food} < 1$), suitable for beta regression.

Usage

```
FoodExpenditure
```

Format

A data frame with 38 rows and 3 variables:

food numeric: proportion of household income spent on food

income numeric: household income (in thousands of dollars)

persons numeric: number of persons living in the household

Source

Griffiths, W. E., Hill, R. C., & Judge, G. G. (1993). *Learning and Practicing Econometrics*. Wiley.

Examples

```
data(FoodExpenditure)
str(FoodExpenditure)
```

GasolineYield *Gasoline Yield Data from Prater (1956)*

Description

A dataset containing 32 observations on gasoline yield under different experimental conditions. The response variable `yield` is a proportion (0 < yield < 1), making it suitable for beta regression.

Usage

```
GasolineYield
```

Format

A data frame with 32 rows and 6 variables:

yield numeric: proportion of crude oil converted to gasoline

batch factor: 10 unique batches of crude oil

temp numeric: temperature (Fahrenheit)

gravity numeric: crude oil gravity

pressure numeric: pressure

temp10 numeric: temperature (scaled)

Source

Prater (1956), as cited in Ferrari and Cribari-Neto (2004) *Beta Regression for Modelling Rates and Proportions* https://www.jstor.org/stable/4110074

Examples

```
data(GasolineYield, package = "betareg")
str(GasolineYield)
```

ReadingSkills *Reading Skills Data*

Description

A dataset from Smithson and Verkuilen (2006) on reading accuracy, dyslexia status, and IQ scores. The response variable `accuracy` is a proportion (0 < accuracy < 1), suitable for beta regression.

Usage

```
ReadingSkills
```

Format

A data frame with 44 rows and 4 variables:

accuracy numeric: proportion of correct responses in a reading task

accuracy1 numeric: transformed accuracy measure

dyslexia factor: dyslexia status (levels: "yes", "no")

iq numeric: IQ score

Source

Smithson, M. & Verkuilen, J. (2006). *A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables*. https://psycnet.apa.org/doi/10.1037/1082-989X.11.1.54

Examples

```
data(ReadingSkills)
str(ReadingSkills)
```

Index