

# Package ‘SenTinMixt’

January 20, 2025

**Title** Parsimonious Mixtures of MSEN and MTIN Distributions

**Version** 1.0.0

**Description** Implements parsimonious mixtures of MSEN and MTIN distributions via expectation-maximization based algorithms for model-based clustering. For each mixture component, parsimony is reached via the eigen-decomposition of the scale matrices and by imposing a constraint on the tailedness parameter. This produces a family of 28 parsimonious mixture models for each distribution.

**License** GPL ( $\geq 3$ )

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

**Imports** doSNOW, foreach, snow, TSdist, tidyr, data.table, expint,  
zipfR, mclust, rlist, withr

**Depends** R ( $\geq 2.10$ )

**NeedsCompilation** no

**Author** Salvatore D. Tomarchio [aut, cre],  
Bagnato Luca [aut],  
Antonio Punzo [aut]

**Maintainer** Salvatore D. Tomarchio <daniele.tomarchio@unict.it>

**Repository** CRAN

**Date/Publication** 2021-10-20 06:50:02 UTC

## Contents

AIS	2
dmsen	3
dmtin	4
Hawks	5
Mixt.fit	5
Mixt.fit.init	7
rMixt	8
rmsen	9
rmtin	10

---

AIS

*Australian institute of sport data*

---

### Description

A dataset containing biometrical measurements for two categories of athletes collected at the Australian Institute of Sport.

### Usage

```
data(AIS)
```

### Format

A matrix with 202 observations on the following variables:

**Sex** 0 = Male or 1 = Female.

**Ht** Height (in cm).

**LBM** Lean body mass (in Kg).

**RCC** Red cell count.

**Hc** Hematocrit.

**Hg** Hemoglobin.

**SSF** Sum of skin folds.

**Bfat** Body fat percentage.

### Source

This dataset is a subset of the `ais` dataset contained in the `alr4` R package.

### References

Weisberg Sanford (2018). `alr4`: Data to Accompany Applied Linear Regression 4th Edition. <https://CRAN.R-project.org/package=alr4>.

---

dmsen	<i>Density of a MSEN distribution</i>
-------	---------------------------------------

---

### Description

Density of a MSEN distribution

### Usage

```
dmsen(x, mu = rep(0, d), Sigma, theta = Inf, formula = "direct")
```

### Arguments

x	A data matrix with n rows and d columns, being n the number of data points and d the data the dimensionality.
mu	A vector of length d representing the mean value.
Sigma	A symmetric positive-definite matrix representing the scale matrix of the distribution.
theta	A number greater than 0 indicating the tailedness parameter.
formula	Method used to calculate the density: "direct", "indirect", "series".

### Value

The value(s) of the density in x

### References

Punzo A., and Bagnato L. (2020). Allometric analysis using the multivariate shifted exponential normal distribution. *Biometrical Journal*, **62**(6), 1525-1543.

### Examples

```
d <- 3
x <- matrix(rnorm(d*2), 2, d)
dmsen(x, mu = rep(0,d), Sigma = diag(d), theta = 0.4, formula = "direct")
```

---

dmtin	<i>Density of a MTIN distribution</i>
-------	---------------------------------------

---

### Description

Density of a MTIN distribution

### Usage

```
dmtin(x, mu = rep(0, d), Sigma, theta = 0.01, formula = "direct")
```

### Arguments

x	A data matrix with n rows and d columns, being n the number of data points and d the data the dimensionality.
mu	A vector of length d representing the mean value.
Sigma	A symmetric positive-definite matrix representing the scale matrix of the distribution.
theta	A number greater than 0 indicating the tailedness parameter.
formula	Method used to calculate the density: "direct", "indirect", "series".

### Value

The value(s) of the density in x

### References

Punzo A., and Bagnato L. (2021). The multivariate tail-inflated normal distribution and its application in finance. *Journal of Statistical Computation and Simulation*, **91**(1), 1-36.

### Examples

```
d <- 3
x <- matrix(rnorm(d*2), 2, d)
dmtin(x, mu = rep(0,d), Sigma = diag(d), theta = 0.9, formula = "direct")
```

---

Hawks

*Measurements on Two Hawk Species*

---

### Description

A dataset containing size-related measurements for two different Hawk species. Each species is further categorized by sex.

### Usage

```
data(Hawks)
```

### Format

A matrix with 323 observations on the following variables:

**Class** 1 = Male CH hawks, 2 = Male SS hawks, 3 = Female CH hawks or 4 = Female SS hawks

**Wing** Length (in mm) of primary wing feather from tip to wrist it attaches to.

**Weight** Body weight (in gm).

**Tail** Measurement (in mm) related to the length of the tail.

### Source

This dataset is a subset of the Hawks dataset contained in the Stat2Data R package.

### References

Cannon et al. (2019). Stat2Data: Datasets for Stat2. <https://CRAN.R-project.org/package=Stat2Data>.

---

Mixt.fit

*Fitting for parsimonious mixtures of MSEN or MTIN distributions*

---

### Description

Fits, by using EM-based algorithms, parsimonious mixtures of MSEN or MTIN distributions to the given data. Parallel computing is implemented and highly recommended for a faster model fitting. The Bayesian information criterion (BIC) and the integrated completed likelihood (ICL) are used to select the best fitting models according to each information criterion.

**Usage**

```
Mixt.fit(
  X,
  k = 1:3,
  init.par = NULL,
  cov.model = "all",
  theta.model = "all",
  density,
  ncores = 1,
  verbose = FALSE,
  ret.all = FALSE
)
```

**Arguments**

<code>X</code>	A data matrix with <code>n</code> rows and <code>d</code> columns, being <code>n</code> the number of data points and <code>d</code> the data the dimensionality.
<code>k</code>	An integer or a vector indicating the number of groups of the models to be estimated.
<code>init.par</code>	The initial values for starting the algorithms, as produced by the <code>Mixt.fit.init()</code> function.
<code>cov.model</code>	A character vector indicating the parsimonious structure of the scale matrices. Possible values are: "EII", "VII", "EEI", "VEI", "EVI", "VVI", "EEE", "VEE", "EVE", "EEV", "VVE", "VEV", "EVV", "VVV" or "all". When "all" is used, all of the 14 parsimonious structures are considered.
<code>theta.model</code>	A character vector indicating the parsimonious structure of the tailedness parameters. Possible values are: "E", "V" or "all". When "all" is used, both parsimonious structures are considered.
<code>density</code>	A character indicating the density of the mixture components. Possible values are: "MSEN" or "MTIN".
<code>ncores</code>	A positive integer indicating the number of cores used for running in parallel.
<code>verbose</code>	A logical indicating whether the running output should be displayed.
<code>ret.all</code>	A logical indicating whether to report the results of all the models or only those of the best models according to BIC and ICL.

**Value**

A list with the following elements:

<code>all.models</code>	The results related to the all the fitted models (only when <code>ret.all=TRUE</code> ).
<code>BicWin</code>	The best fitting model according to the BIC.
<code>IclWin</code>	The best fitting model according to the ICL.
<code>Summary</code>	A quick table showing summary results for the best fitting models according to BIC and ICL.

**Examples**

```

set.seed(1234)
n <- 50
k <- 2
Pi <- c(0.5, 0.5)
mu <- matrix(c(0, 0, 4, 5), 2, 2)
cov.model <- "EEE"
lambda <- c(0.5, 0.5)
delta <- c(0.7, 0.7)
gamma <- c(2.62, 2.62)
theta <- c(0.1, 0.1)
density <- "MSEN"
data <- rMixt(n, k, Pi, mu, cov.model, lambda, delta, gamma, theta, density)

X <- data$X
nstartR <- 1
init.par <- Mixt.fit.init(X, k, density, nstartR)

theta.model <- "E"
res <- Mixt.fit(X, k, init.par, cov.model, theta.model, density)

```

---

Mixt.fit.init

*Initialization for the EM-based algorithms*


---

**Description**

Runs the initialization of the EM-based algorithms used for fitting parsimonious mixtures of MSEN or MTIN distributions. Parallel computing is implemented and highly recommended for a faster calculation.

**Usage**

```
Mixt.fit.init(X, k = 1:3, density, nstartR = 100, ncores = 1, verbose = FALSE)
```

**Arguments**

X	A data matrix with n rows and d columns, being n the number of data points and d the data the dimensionality.
k	An integer or a vector indicating the number of groups of the models.
density	A character indicating the density of the mixture components. Possible values are: "MSEN" or "MTIN".
nstartR	An integer specifying the number of random starts to be considered.
ncores	A positive integer indicating the number of cores used for running in parallel.
verbose	A logical indicating whether the running output should be displayed.

**Value**

init A list of objects to be used by the `Mixt.fit()` function.

## Examples

```

set.seed(1234)
n <- 50
k <- 2
Pi <- c(0.5, 0.5)
mu <- matrix(c(0, 0, 4, 5), 2, 2)
cov.model <- "EEE"
lambda <- c(0.5, 0.5)
delta <- c(0.7, 0.7)
gamma <- c(2.62, 2.62)
theta <- c(0.1, 0.1)
density <- "MSEN"
data <- rMixt(n, k, Pi, mu, cov.model, lambda, delta, gamma, theta, density)

X <- data$X
nstartR <- 1
init.par <- Mixt.fit.init(X, k, density, nstartR)

```

---

rMixt

*Random number generation for bidimensional parsimonious mixtures of MSEN or MTIN distributions*

---

## Description

Random number generation for bidimensional parsimonious mixtures of MSEN or MTIN distributions

## Usage

```
rMixt(n, k, Pi, mu, cov.model, lambda, delta, gamma, theta, density)
```

## Arguments

n	An integer specifying the number of data points to be simulated.
k	An integer indicating the number of groups in the data.
Pi	A vector of length k representing the probability of belonging to the k groups for each data point.
mu	A matrix of means with 2 rows and k columns.
cov.model	A character indicating the parsimonious structure of the scale matrices. Possible values are: "EII", "VII", "EII", "VEI", "EVI", "VVI", "EEE", "VEE", "EVE", "EEV", "VVE", "VEV", "EVV" or "VVV".
lambda	A numeric vector of length k, related to the scale matrices (see Punzo et al., 2016), which determines the volumes of the mixture components. Each element must be greater than 0. Required for all the parsimonious structures.



delta	A numeric vector of length k, related to the scale matrices (see Punzo et al., 2016), which determines the shapes of the mixture components. Each element must be between 0 and 1. Required for all the parsimonious structures, with the exclusion of "EII" and "VII".
gamma	A numeric vector of length k, related to the scale matrices (see Punzo et al., 2016), which determines the orientation of the mixture components. Each element represents an angle expressed in radian unit. Required for the "EEE", "VEE", "EVE", "EEV", "VVE", "VEV", "EVV" or "VVV" parsimonious structures.
theta	A vector of length k representing the tailedness parameters.
density	A character indicating the density of the mixture components. Possible values are: "MSEN" or "MTIN".

### Value

A list with the following elements:

X	A data matrix with n rows and 2 columns.
Sigma	An array of dimension 2 x 2 x k for the generated scale matrices.
Size	The size of each generated group.

### References

Punzo A., Browne R. and McNicholas P.D. (2016). Hypothesis Testing for Mixture Model Selection. *Journal of Statistical Computation and Simulation*, **86**(14), 2797-2818.

### Examples

```
n <- 50
k <- 2
Pi <- c(0.5, 0.5)
mu <- matrix(c(0, 0, 4, 5), 2, 2)
cov.model <- "EEE"
lambda <- c(0.5, 0.5)
delta <- c(0.7, 0.7)
gamma <- c(2.62, 2.62)
theta <- c(0.1, 0.1)
density <- "MSEN"
data <- rMixt(n, k, Pi, mu, cov.model, lambda, delta, gamma, theta, density)
```

---

rmsen

*Random number generation for the MSEN distribution*

---

### Description

Random number generation for the MSEN distribution

**Usage**

```
rmsen(n, mu = rep(0, d), Sigma, theta = Inf)
```

**Arguments**

**n** An integer specifying the number of data points to be simulated.

**mu** A vector of length  $d$ , where  $d$  is the dimensionality, representing the mean value.

**Sigma** A symmetric positive-definite matrix representing the scale matrix of the distribution.

**theta** A number greater than 0 indicating the tailedness parameter.

**Value**

A list with the following elements:

**X** A data matrix with  $n$  rows and  $d$  columns.

**w** A vector of weights of dimension  $n$ .

**References**

Punzo A., and Bagnato L. (2020). Allometric analysis using the multivariate shifted exponential normal distribution. *Biometrical Journal*, **62**(6), 1525-1543.

**Examples**

```
d <- 3
rmsen(10, mu = rep(0, d), Sigma = diag(d), theta = 0.3)
```

---

rmtin

*Random number generation for the MTIN distribution*

---

**Description**

Random number generation for the MTIN distribution

**Usage**

```
rmtin(n, mu = rep(0, d), Sigma, theta = 0.01)
```

**Arguments**

**n** An integer specifying the number of data points to be simulated.

**mu** A vector of length  $d$ , where  $d$  is the dimensionality, representing the mean value.

**Sigma** A symmetric positive-definite matrix representing the scale matrix of the distribution.

**theta** A number between 0 and 1 indicating the tailedness parameter.

**Value**

A list with the following elements:

- X                    A data matrix with n rows and d columns.
- w                    A vector of weights of dimension n.

**References**

Punzo A., and Bagnato L. (2021). The multivariate tail-inflated normal distribution and its application in finance. *Journal of Statistical Computation and Simulation*, **91**(1), 1-36.

**Examples**

```
d <- 3
rmtin(10, mu = rep(0, d), Sigma = diag(d), theta = 0.9)
```

# Index

## \* datasets

AIS, 2

Hawks, 5

AIS, 2

dmsen, 3

dmtin, 4

Hawks, 5

Mixt.fit, 5

Mixt.fit.init, 7

rMixt, 8

rmsen, 9

rmtin, 10