Package: cmgnd (via r-universe)

November 20, 2026

Type Package

Title Constrained Mixture of Generalized Normal Distributions

Version 0.1.1

Description The 'cmgnd' implements the constrained mixture of generalized normal distributions model, a flexible statistical framework for modelling univariate data exhibiting non-normal features such as skewness, multi-modality, and heavy tails. By imposing constraints on model parameters, the 'cmgnd' reduces estimation complexity while maintaining high descriptive power, offering an efficient solution in the presence of distributional irregularities. For more details see Duttilo and Gattone (2025) <doi:10.1007/s00180-025-01638-x> and Duttilo et al (2025) <doi:10.48550/arXiv.2506.03285>.

Depends R (>= 4.0.0)

Imports RcppAlgos, lubridate, stats, utils, gnorm, purrr, ggplot2

Encoding UTF-8

RoxygenNote 7.2.3

License GPL (>= 3)

URL https://github.com/pierdutt/cmgnd

BugReports https://github.com/pierdutt/cmgnd/issues

Config/pak/sysreqs libgmp3-dev

Repository https://pierdutt.r-universe.dev

Date/Publication 2025-07-27 22:22:33 UTC

RemoteUrl https://github.com/pierdutt/cmgnd

RemoteRef HEAD

RemoteSha ba0015b3f1b59b6f2404cc78fee176cf6145400e

Contents

cmgnd 2
dcmgnd 5
dgnd 6
hist_cmgnd 7
moments_cmgnd 8
plot_cmgnd 9
sim_cmgnd 10

Index **11**

---

cmgnd *cmgnd: Function for Clustering using Constrained Mixtures of Generalized Normal Distributions*

Description

Fits univariate constrained mixture of generalized normal distribution models by imposing mixture partitions. Models are estimated by the ECM algorithm initialized by k-means.

Usage

```
cmgnd(
  x,
  K = 2,
  Cmu = rep(0, K),
  Csigma = rep(0, K),
  Cnu = rep(0, K),
  nstart = 50,
  theta = FALSE,
  nustart = rep(2, K),
  nustartype = "random",
  gauss = FALSE,
  laplace = FALSE,
  scale = FALSE,
  eps = 10^-4,
  maxit = 999,
  verbose = TRUE,
  sigbound = c(0.1, 5),
  sr = "like",
  eta = 0.5,
  seed = 12345,
  seed.nstart = seq(1:nstart)
)
```

Arguments

x	A numeric vector of observations.
K	An integer specifying the number of mixture components to fit. Default is 2.
Cmu	A binary vector indicating mixture components for the location parameter. The k-th element is set to 1 if the k-th mixture component belongs to the Cr partition, and 0 otherwise. Default is c(0,0), indicating no mixture partition with K=2.
Csigma	A binary vector indicating mixture components for the scale parameter. The k-th element is set to 1 if the k-th mixture component belongs to the Cr partition, and 0 otherwise. Default is c(0,0), indicating no mixture partition with K=2.
Cnu	A binary vector indicating mixture components for the shape parameter. The k-th element is set to 1 if the k-th mixture component belongs to the Cr partition, and 0 otherwise. Default is c(0,0), indicating no mixture partition with K=2.
nstart	An integer specifying the number of starting points for the shape parameter. Default is 10.
theta	A parameter matrix used to initialize the estimation for the first starting point.
nustart	A numeric vector containing the starting values for the shape parameter nu. Default is c(2,2) for K=2.
nustartype	A character string indicating whether the initialization of nu should be "random", around the values in nustart, or "exact", using the exact values in nustart.
gauss	A logical value indicating if the algorithm should use the Gaussian distribution. Default is FALSE.
laplace	A logical value indicating if the algorithm should use the Laplace distribution. Default is FALSE.
scale	A logical value indicating whether the function should scale the data. Default is TRUE.
eps	A numeric value specifying the tolerance level of the ECM algorithm.
maxit	An integer specifying the maximum number of iterations.
verbose	A logical value indicating whether to display running output. Default is TRUE.
sigbound	A numeric vector of length two specifying the lower and upper bounds for resetting the sigma estimates. Default value is c(.01,5).
sr	A character string specifying the type of convergence criterion to use. The default is "like", but "parameter" can be used for likelihood-based convergence.
eta	A numeric value specifying the tolerance level for the likelihood-based convergence. Default value is .5.
seed	Optional integer used to set the random seed via set.seed(). The default is 12345. If NULL, no seed is set and results may vary between runs.
seed.nstart	Optional numeric vector used to set the random seed via set.seed() during the initialisation of the nstart random posterior probabilities. The default is seq_len(nstart). If NULL, no seed is set and results may vary between runs.

Details

The constrained mixture of generalized normal distributions (CMGND) model is an advanced statistical tool designed for analyzing univariate data characterized by non-normal features such as asymmetry, multi-modality, leptokurtosis, and heavy tails. This model extends the mixture of generalized normal distributions (MGND) by incorporating constraints on the parameters, thereby reducing the number of parameters to be estimated and improving model performance. The CMGND model is defined by the following components:

$$f(x|\theta) = \sum_{k=1}^{K} \pi_k f_k(x|\mu_k, \sigma_k, \nu_k)$$

where: π_k are the mixture weights, satisfying $0 < \pi_k < 1$ and $\sum_{k=1}^{K} \pi_k = 1$. $f_k(x|\mu_k, \sigma_k, \nu_k)$ is the Generalized Normal Distribution for the k-th component with mean μ_k, scale σ_k, and shape parameter ν_k.

The parameter space can be constrained by imposing equality constraints such as $\mu_k = \mu_r$, $\sigma_k = \sigma_r$, and/or $\nu_k = \nu_r$ for all $k \in C_r$, where C_r is a partition of the set $\{1, 2, \ldots, K\}$.

The $k \in C_r$ partition for each parameter can be specified by the binary vectors `Cmu`, `Csigma` and `Cnu`.

Value

`ll`	The log-likelihood corresponding to the estimated model.
`nobs`	Number of observations.
`parameters`	Data frame of the estimated parameters.
`ic`	Data frame of information criteria. AIC, BIC, HQIC and EDC are returned.
`res`	Matrix of posterior probabilities or responsibilities.
`clus`	Vector of group classifications.
`op_it`	List containing three integers: `permstart` the optimal starting value of the permutation of k-means solutions; `startnu` the optimal starting value of the shape parameter; `iter` number of iterations.
`cputime`	A numeric value indicating the cpu time employed.
`info`	List containing a few of the original user inputs, for use by other dedicated functions of the `cmgnd` class.

References

Bazi, Y., Bruzzone, L., and Melgani, F. (2006). Image thresholding based on the em algorithm and the generalized gaussian distribution. Pattern Recognition, 40(2), pp 619–634.

Wen, L., Qiu, Y., Wang, M., Yin, J., and Chen, P. (2022). Numerical characteristics and parameter estimation of finite mixed generalized normal distribution. Communications in Statistics - Simulation and Computation, 51(7), pp 3596–3620.

Duttilo, P. (2024). Modelling financial returns with mixtures of generalized normal distributions. PhD Thesis, University "G. d'Annunzio" of Chieti-Pescara, pp. 1-166, doi:10.48550/arXiv.2411.11847

Duttilo, P. and Gattone, S.A. (2025). Enhancing parameter estimation in finite mixture of generalized normal distributions, Computational Statistics, pp. 1-28, doi:10.1007/s0018002501638x

Duttilo, P., Gattone, S.A., and Kume A. (2025). Constrained mixtures of generalized normal distributions, pp. 1-34, doi:10.48550/arXiv.2506.03285

Examples

```
# Old Faithful dataset
x=faithful$eruptions
# Unconstrained model estimation
Cmu <- c(0, 0)
Csigma <- c(0, 0)
Cnu <- c(0, 0)
model_unc <- cmgnd(x, nstart = 2, K = 2, Cmu, Csigma, Cnu)
model_unc$parameters
plot_cmgnd(x, model_unc)
# Constrained model estimation with common scale parameters
Csigma <- c(1, 1)
model_con <- cmgnd(x, nstart = 2, K =2, Cmu, Csigma, Cnu)
model_con$parameters
plot_cmgnd(x, model_con)
```

---

dcmgnd *Marginal Density Estimation for CMGND Models*

Description

This function estimates the marginal density for univariate constrained mixture of generalized normal distribution (CMGND) models.

Usage

```
dcmgnd(x, parameters)
```

Arguments

`x`	A numeric vector representing the observed data points.
`parameters`	A matrix or data.frame containing the parameters of the CMGND model. This can also be an object returned from the ‘cmgnd()‘ function, representing a previously estimated CMGND model.

Details

The function computes the marginal density based on the provided parameters of the CMGND model. It can handle both newly supplied parameters or those extracted from an existing CMGND model object.

Value

A vector of density estimates corresponding to the input data ‘x‘.

See Also

‘cmgnd()‘ for estimating the model parameters.

---

dgnd *The Generalized Normal Distribution (GND)*

Description

Density function for the GND with location parameter mu, scale parameter sigma and shape parameter nu.

Usage

```
dgnd(x, mu = 0, sigma = 1, nu = 2)
```

Arguments

x	A numeric vector of observations.
mu	A numeric value indicating the location parameter μ.
sigma	A numeric value indicating the scale parameter σ.
nu	A numeric value indicating the shape parameter ν.

Details

If mu, sigma and nu are not specified they assume the default values of 0, 1 and 2, respectively. The GND distribution has density

$$f_{GND}(x|\mu,\sigma,\nu) = \frac{\nu}{2\sigma\Gamma(1\,/\,\nu)}\exp\left\{-\left|\frac{x-\mu}{\sigma}\right|^{\nu}\right\}.$$

The shape parameter ν controls both the peakedness and tail weights. If $\nu = 1$ the GND reduces to the Laplace distribution and if $\nu = 2$ it coincides with the normal distribution. It is noticed that $1 < \nu < 2$ yields an intermediate distribution between the normal and the Laplace distribution. As limit cases, for $\nu \to \infty$ the distribution tends to a uniform distribution, while for $\nu \to 0$ it will be impulsive.

Value

dgnd returns the density.

References

Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, $32(7)$: $685\tilde{}694$.

`hist_cmgnd` *Plot Marginal and Mixture Component Densities of the CMGND Model*

Description

This function generates a plot displaying both the marginal density and individual mixture component densities for univariate constrained mixture of generalized normal distribution (CMGND) models. It visually represents how the different components of the mixture model contribute to the overall density.

Usage

```
hist_cmgnd(x, parameters, bins = 80)
```

Arguments

`x`	A numeric vector representing the observed data points.
`parameters`	A matrix or data.frame containing the parameters of the CMGND model.
`bins`	Number of bins. Defaults to 80. Alternatively, this can be an object returned from the ‘cmgnd()‘ function, representing an estimated CMGND model.

Details

The function plots the overall (marginal) density curve for the CMGND model, as well as the density curves of each mixture component. This visualization helps in understanding how each component contributes to the model and provides insights into the data distribution.

Value

A plot illustrating the marginal density along with the densities of the individual mixture components for the given data ‘x‘.

See Also

‘cmgnd()‘ for estimating the model parameters.

`moments_cmgnd` *Compute the First Four Moments of the CMGND Marginal Distribution*

Description

Computes the mean, variance, skewness, and kurtosis of the marginal distribution of a univariate Constrained Mixture of Generalized Normal Distributions (CMGND) model, given the model parameters.

Usage

```
moments_cmgnd(parameters)
```

Arguments

`parameters` A matrix or data frame where each row corresponds to a component of the mixture. Columns must be ordered as follows:

1. Mixing proportions π_k
2. Component means μ_k
3. Scale parameters σ_k
4. Shape parameters ν_k

Details

The function assumes that the parameters define a valid CMGND model and uses analytical expressions to compute the first four moments of the marginal distribution. The shape parameter ν_k governs the kurtosis of each component.

Value

A named list with the following elements:

mean The marginal mean of the CMGND distribution

var The marginal variance

skew The marginal skewness

kur The marginal kurtosis

See Also

`cmgnd` for estimating CMGND model parameters.

plot_cmgnd *Plot Marginal and Mixture Component Densities of the CMGND Model*

Description

This function generates a plot displaying both the marginal density and individual mixture component densities for univariate constrained mixture of generalized normal distribution (CMGND) models. It visually represents how the different components of the mixture model contribute to the overall density.

Usage

```
plot_cmgnd(x, parameters, model = "")
```

Arguments

x	A numeric vector representing the observed data points.
parameters	A matrix or data.frame containing the parameters of the CMGND model.
model	A character indicating the model type name. Default value model="". Alternatively, this can be an object returned from the 'cmgnd()' function, representing an estimated CMGND model.

Details

The function plots the overall (marginal) density curve for the CMGND model, as well as the density curves of each mixture component. This visualization helps in understanding how each component contributes to the model and provides insights into the data distribution.

Value

A plot illustrating the marginal density along with the densities of the individual mixture components for the given data 'x'.

See Also

'cmgnd()' for estimating the model parameters.

`sim_cmgnd` *sim_cmgnd: Function to Simulate Univariate Constrained Mixtures of Generalized Normal Distributions*

Description

Simulate univariate constrained mixture of generalized normal distribution models. Remeber to set the set.seed() before the function sim_cmgnd().

Usage

```
sim_cmgnd(
  n = 1000,
  pi = rep(0.5, 2),
  mu = c(1, 5),
  sigma = c(1, 1),
  nu = c(2, 2)
)
```

Arguments

`n`	A numeric value indicating the total number of observations to simulate.
`pi`	A numeric vector of the mixture weights π_k.
`mu`	A numeric vector of the location parameter μ_k.
`sigma`	A numeric vector of the scale parameter σ_k.
`nu`	A numeric vector of the shape parameter ν_k.

Value

`sim_data`	The simulated data.
`sim_clus`	The cluster indication of simulated data.

Index

cmgnd, 2, *8*

dcmgnd, 5
dgnd, 6

hist_cmgnd, 7

moments_cmgnd, 8

plot_cmgnd, 9

sim_cmgnd, 10