Package 'CDGHMM'

November 1, 2024

Type Package

Title Hidden Markov Models for Multivariate Panel Data

Version 0.1.1

Date 2024-11-01

Maintainer Mackenzie R. Neal <nealm6@mcmaster.ca>

Description Estimates hidden Markov models from the family of Cholesky-decomposed Gaussian hidden Markov models (CDGHMM) under various missingness schemes. This family improves upon estimation of traditional Gaussian HMMs by introducing parsimony, as well as, controlling for dropped out observations and non-random missingness. See Neal, Sochaniwsky and McNicholas (2024) <DOI:10.1007/s11222-024-10462-0>.

License GPL (>= 2)

Imports MASS, mvtnorm, ramify, cluster

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Author Mackenzie R. Neal [aut, cre] (<https://orcid.org/0009-0004-4619-3247>),
Alexa A. Sochaniwsky [aut],
Paul D. McNicholas [aut] (<https://orcid.org/0000-0002-2482-523X>)

Depends R (>= 3.5.0)

Date/Publication 2024-11-01 17:20:05 UTC

Contents

cdghmm *Hidden Markov Models for Multivariate Panel Data*

Description

Estimates hidden Markov models from the CDGHMM family under various missingness schemes.

Usage

```
cdghmm(x,m,id,mu=NULL,sigma=NULL,gamma=NULL,delta=NULL,alpha=NULL,beta=NULL,
                        maxiter=10000,tol=1e-6,type="s",covtype="VVA")
```

Arguments

x	Data frame or matrix to perform variable selection on
m	Number to indicate the number of states to fit.
id	A vector of indicators to indicate observational unit.
mu	An m-length list of matrices to be used as an initial estimate for mu. If no initial mu is provided the algorithm will initialize via k-means.
sigma	An m-length list of matrices to be used as an initial estimate for sigma. If no initial sigma is provided the algorithm will initialize.
gamma	A mxm matrix to be used as an initial estimate for gamma. If no initial gamma is provided the algorithm will initialize.
delta	A vector to be used as an initial estimate for delta. If no initial delta is provided the algorithm will initialize.
alpha	A mxpxt array to be used as an initial estimate for alpha. If no initial alpha is provided the algorithm will initialize.
beta	A mxpxt array to be used as an initial estimate for beta. If no initial beta is provided the algorithm will initialize.
maxiter	A number to indicate the maximum number of iterations allowed, default is 10000.
tol	A number to indicate the tolerance value, default is 1e-6.
type	A character to indicate which type of missingness mechanism to use. The allowed values are: "mar" (missing at random), "s" (3.2.1 in cited paper), "sv" (3.2.2), "st" (3.2.3 model without beta), "svt" (3.2.4 model without beta), "st2" (3.2.3 model with beta), "svt2" (3.2.4 model with beta). The default is type="s".
covtype	A string to indicate which covariance estimate to use. The allowed values are: EEA, VVA, VEA, EVA, VVI, VEI, EVI, EEI. The default is covtype="VVA".

Value

mu	Estimated mean matrices.
sigma	Estimated covariance matrices.
gamma	Estimated gamma matrix.
delta	Estimated delta vector.
alpha	Estimated alpha missingness parameters.
beta	Estimated beta missingness parameters.
llk	Estimated log-likelihood.
AIC	The value of the Akaike information criterion.
BIC	The value of the Bayes information criterion.
ICL	The value of the integrated completed likelihood.
Avg_Silouette	The value of the average silhouette score.
probs	A matrix whose entries correspond to the probability of belonging to a state.
states	Estimated states via map(probs).
mod	The CDGHMM family member fit.

Author(s)

Mackenzie R. Neal, Alexa A. Sochaniwsky, Paul D. McNicholas

References

See `citation("CDGHMM")`.

Examples

```
data("simulated_data")
id=simulated_data$V5
x <- simulated_data[,1:4]
EEI_mod=cdghmm(x,2,id=id,covtype="EEI",tol=1e-4)
table(simulated_data$V7,EEI_mod$states)
```

---

simulated_data *Simulated data from Simulation 1 in cited paper.*

Description

Simulated data with two groups, used to illustrate `cdghmm`

Usage

```
data(simulated_data)
```

Source

These data were simulated using R.

Index