Package 'COST'

January 20, 2025

Type Package

Title Copula-Based Semiparametric Models for Spatio-Temporal Data

Version 0.1.0

Author Yanlin Tang, Huixia Judy Wang

Maintainer Yanlin Tang <yanlintang2018@163.com>

Description Parameter estimation, one-step ahead forecast and new location prediction methods for spatio-temporal data.

Depends copula,mvtnorm

License GPL

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

ByteCompile yes

NeedsCompilation no

Repository CRAN

Date/Publication 2019-01-04 11:00:24 UTC

Contents

`Data.COST` *Data Generation*

Description

Generating data from COST DGP, assuming Markov process of order one

Usage

```
Data.COST(n,n.total,seed1,coord,par.t)
```

Arguments

`n`	number of time points for parameter estimation
`n.total`	number of total time points, with a burning sequence
`seed1`	random seed to generate a data set, for reproducibility
`coord`	coordinates of the locations
`par.t`	the true copula parameters

Value

`Y.all`	data from all locations and time points, may include data at time point n+1, or data from new locations
`mean.true`	true conditional mean of observed locations at time point n+1

Author(s)

Yanlin Tang, Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

Examples

```
library(COST)
n = 500
n.total = 1001
seed1 = 22222
coord = cbind(rep(c(1,3,5)/6,each=3),rep(c(1,3,5)/6,3))
par.t = c(0,1,1,0.5,1.5,100)
dat = Data.COST(n,n.total,seed1,coord,par.t)
#it returns a data set with dimension 501*9
```

`example.forecast` *example for one-step ahead forecast*

Description

Example for one-step ahead forecast for Gaussian Process and our COST method with Gaussian and t copulas, where the data are generated from COST DGP, where the parameters are assumed to be known; the parameters can be obtained by the “optim" function. Assuming that data are observed at d=9 locations, and n+1 time points, where the last time point is for validation.

Usage

```
example.forecast(n,n.total,seed1)
```

Arguments

`n`	number of time points for parameter estimation
`n.total`	number of total time points, with a burning sequence
`seed1`	random seed to generate a data set, for reproducibility

Value

`COST.t.fore.ECP`
: a vector of length d, with value 1 or 0, 1 means the verifying value from the corresponding location lies in the 95% forecast interval, 0 means not

`COST.t.fore.ML`
: a vector of length d, each element is the length of forecast interval of the corresponding location

`COST.t.fore.rank`
: multivariate rank of the verifying vector by t copula

`COST.G.fore.ECP`
: same as COST.t.fore.ECP

`COST.G.fore.ML`
: same as COST.t.fore.ML

`COST.G.fore.rank`
: multivariate rank of the verifying vector by Gaussian copula

`GP.fore.ECP`
: same as COST.t.fore.ECP

`GP.fore.ML`
: same as COST.t.fore.ML

`GP.fore.rank`
: multivariate rank of the verifying vector by Gaussian process method

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

Examples

```
library(COST)
#settings
seed1 = 2222222
n.total = 101 #number of total time points, including the burning sequence
n = 50 #number of time points we observed
example.forecast(n,n.total,seed1)
#OUTPUTS

# $COST.t.fore.ECP #whether the forecast interval includes the true value at n+1
# [1] 1 1 1 1 1 1 1 1 1
#
# $COST.t.fore.ML #length of the forecast interval
# [1] 0.7036 4.1318 4.8749 2.7615 3.7398 5.8186 4.4532 4.9251 6.3757
#
# $COST.t.fore.rank #multivariate rank
# [1] 162
#
#
# $COST.G.fore.ECP #whether the forecast interval includes the true value at n+1
# [1] 1 1 1 1 1 1 1 1 1
#
# $COST.G.fore.ML #length of the forecast interval
# [1]  0.7035 4.1316 4.8656 2.7611 3.7388 5.7913 4.4458 4.9036 6.3727
#
# $COST.G.fore.rank #multivariate rank
# [1] 186
#

# $GP.fore.ECP #whether the forecast interval includes the true value at n+1
# [1] 1 0 0 1 1 1 1 1 1
#
# $GP.fore.ML #length of the forecast interval
# [1] 0.4879 2.0449 3.4436 2.2107 2.9170 4.4537 4.2169 5.5789 7.3689
#
# $GP.fore.rank #multivariate rank
# [1] 17
```

example.prediction *example for new location prediction*

Description

Example for new location prediction, Gaussian process method, and our COST method with Gaussian and t copulas, where the parameters are assumed to be known; the parameters can be obtained by the “optim" function. Data are generated at 13 locations and n time points, and assume that 9 locations are observed, and 4 new locations need prediction at time n, conditional on 9 locations at time points n-1 and n.

Usage

```
example.prediction(n,n.total,seed1)
```

Arguments

`n`	number of time points for parameter estimation
`n.total`	number of total time points, with a burning sequence
`seed1`	random seed to generate a data set, for reproducibility

Value

`COST.t.pre.ECP`	a vector of length K=4 (number of new locations), with value 1 or 0, 1 means the verifying value from the corresponding location lies in the 95% prediction interval, 0 means not
`COST.t.pre.ML`	a vector of length K=4, each element is the length of prediction interval of the corresponding location
`COST.t.pre.med.error`	prediction error based on conditional median
`COST.G.pre.ECP`	same as COST.t.pre.ECP
`COST.G.pre.ML`	same as COST.t.pre.ML
`COST.G.pre.med.error`	same as COST.t.pre.med.error
`GP.pre.ECP`	same as COST.t.pre.ECP
`GP.pre.ML`	same as COST.t.pre.ML
`GP.pre.med.error`	same as COST.t.pre.med.error

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

Examples

```
library(COST)
#settings
n.total = 101 #number of total time points, including the burning sequence
n = 50 #number of time points we observed
seed1 = 22222
example.prediction(n,n.total,seed1)

#OUTPUTS

# $COST.t.pre.ECP #whether the prediction interval includes the true value, time point n
# [1] 1 1 1 1
#
# $COST.t.pre.ML #length of the prediction interval
# [1] 1.445576 2.146452 2.260688 2.706681
#
# $COST.t.pre.med.error #point prediction error, using conditional median
# [1]  0.01127162 -0.03222058 -0.22081051  0.57831480
#
# $COST.G.pre.ECP #whether the prediction interval includes the true value, time point n
# [1] 1 1 1 1
#
# $COST.G.pre.ML #length of the prediction interval
# [1] 1.445576 2.432646 2.260688 2.914887
#
# $COST.G.pre.med.error #point prediction error, using conditional median
# [1] 0.01127162 -0.03222058 -0.22081051  0.57831480
#
# $GP.pre.ECP #whether the prediction interval includes the true value, time point n
# [1] 1 1 1 1
#
# $GP.pre.ML #length of the prediction interval
# [1] 0.8345359 1.4096642 1.5948724 2.3419428
#
# $GP.pre.med.error #point prediction error, using conditional median
# [1] 0.09447685 -0.05889409 -0.08923935  0.58494684
```

`Forecasts.CF` *one-step ahead forecast by separate time series analysis*

Description

one-step ahead forecast, analyzing the time series at each location separately with a t copula, including: (i) point forecast, either conditional median or mean; (ii) 95% forecast intervals, which can also be adjusted by the users; (iii) m (m=500 by default) random draws from the conditional distribution for each location, can be used for multivariate rank after combining all the locations together

Usage

```
Forecasts.CF(par,Y,seed1,m)
```

Arguments

`par`	parameters in the copula function
`Y`	observed data
`seed1`	random seed used to generate random draws from the conditional distribution, for reproducibility
`m`	number of random draws to approximate the conditional distribution

Value

`y.qq`	0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each location
`mean.est`	conditional mean estimate for each location
`y.draw.random`	m random draws from the conditional distribution

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`Forecasts.COST.G` *one-step ahead forecast by Gaussian copula*

Description

one-step ahead forecast by Gaussian copula, including: (i) point forecast, either conditional median or mean; (ii) 95% forecast intervals, which can also be adjusted by the users; (iii) m (m=500 by default) random draws from the conditional distribution, can be used for multivariate rank

Usage

```
Forecasts.COST.G(par,Y,s.ob,seed1,m,isotropic)
```

Arguments

`par`	parameters in the copula function
`Y`	observed data
`s.ob`	coordinates of observed locations
`seed1`	random seed used to generate random draws from the conditional distribution, for reproducibility
`m`	number of random draws to approximate the conditional distribution
`isotropic`	indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

`y.qq`	0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each location
`mean.est`	conditional mean estimate for each location
`y.draw.random`	m random draws from the conditional distribution

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`Forecasts.COST.t` *one-step ahead forecast by t copula*

Description

one-step ahead forecast by t copula, including: (i) point forecast, either conditional median or mean; (ii) 95% forecast intervals, which can also be adjusted by the users; (iii) m (m=500 by default) random draws from the conditional distribution, can be used for multivariate rank

Usage

```
Forecasts.COST.t(par,Y,s.ob,seed1,m,isotropic)
```

Arguments

`par`	parameters in the copula function
`Y`	observed data
`s.ob`	coordinates of observed locations
`seed1`	random seed used to generate random draws from the conditional distribution, for reproducibility
`m`	number of random draws to approximate the conditional distribution
`isotropic`	indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

`y.qq`	0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each location
`mean.est`	conditional mean estimate for each location
`y.draw.random`	m random draws from the conditional distribution

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`Forecasts.GP` *one-step ahead forecast by Gaussian process fitting*

Description

one-step ahead forecast by Gaussian process fitting, including: (i) point forecast, either conditional mean; (ii) 95% forecast intervals, which can also be adjusted by the users; (iii) m (m=500 by default) random draws from the conditional distribution, can be used for multivariate rank

Usage

```
Forecasts.GP(par,Y,s.ob,seed1,m,isotropic)
```

Arguments

`par`	parameters in the copula function
`Y`	observed data
`s.ob`	coordinates of observed locations
`seed1`	random seed used to generate random draws from the conditional distribution, for reproducibility
`m`	number of random draws to approximate the conditional distribution
`isotropic`	indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

`y.qq`	0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each location
`mean.est`	conditional mean estimate for each location
`y.draw.random`	m random draws from the conditional distribution

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`location` *Locations of 10 sites*

Description

Locations of 10 sites.

Usage

```
data(location)
```

Format

Locations of 10 sites, 10*2 matrix in Cartesian coordinate system

Source

https://transmission.bpa.gov/business/operations/wind/MetData/default.aspx

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

Examples

```
s.ob = location[-3,2:3]
s.new = location[3,2:3]
```

---

logL.CF *negtive log-likelihood for separate time series analysis*

Description

negtive log-likelihood for separate time series analysis, copula-based semiparametric method from Chen and Fan (2006), assuming t copula for each time series and Markov process of order one, with marginal distribution estimated by espirical CDF, and it is for correlation parameter estimation

Usage

```
logL.CF(par,Yk,dfs)
```

Arguments

`par`	correlation parameter in the t copula function, will be obtained by minimizing the negtive log-likelihood
`Yk`	observed data from k-th location
`dfs`	degrees of freedom for the t copula, obtained from COST method with t copula

Value

the negative log-likelihood

Author(s)

Yanlin Tang and Huixia Judy Wang

References

1.Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. Journal of Econometrics 130, 307–335.\ 2.Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`logL.COST.G` *negtive log-likelihood for Gaussian copula*

Description

gives the negtive log-likelihood of the Gaussian copula, with empirical CDF plugin, and it is for parameter estimation in the correlation matrix

Usage

```
logL.COST.G(par,Y,s.ob)
```

Arguments

`par`	parameters in the copula function, will be obtained by minimizing the negtive log-likelihood
`Y`	the data set from observed locations, used for parameter estimation
`s.ob`	coordinates of observed locations

Value

the negative log-likelihood

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`logL.COST.t` *negtive log-likelihood for t copula*

Description

gives the negtive log-likelihood of the t copula, with empirical CDF plugin, and it is for parameter estimation in the correlation matrix

Usage

```
logL.COST.t(par,Y,s.ob)
```

Arguments

`par`	parameters in the copula function, will be obtained by minimizing the negtive log-likelihood
`Y`	the data set from observed locations, used for parameter estimation
`s.ob`	coordinates of observed locations

Value

the negative log-likelihood

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`logL.GP` *negtive log-likelihood of Gaussian process*

Description

negtive log-likelihood of Gaussian process, with mean vector and variance vector obtained by the empirical version, and it is for parameter estimation in the correlation matrix

Usage

```
logL.GP(par,Y,s.ob)
```

Arguments

`par`	parameters in the copula function, will be obtained by minimizing the negtive log-likelihood
`Y`	the data set from observed locations, used for parameter estimation
`s.ob`	coordinates of observed locations

Value

the negative log-likelihood

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`Predictions.COST.G` *new location prediction by Gaussian copula*

Description

new location prediction by Gaussian copula, where the copula dimension is extended, and the marginal CDF of the new location is estimated by neighboring information; it gives 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n, conditional on observed locations at time n-1 and n; both point and interval predictions are provided

Usage

```
Predictions.COST.G(par,Y,s.ob,s.new,isotropic)
```

Arguments

`par`	parameters in the copula function
`Y`	observed data
`s.ob`	coordinates of observed locations
`s.new`	coordinates of new locations
`isotropic`	indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

Predictions.COST.t *new location prediction by t copula*

Description

new location prediction by t copula, where the copula dimension is extended, and the marginal CDF of the new location is estimated by neighboring information; it gives 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n, conditional on observed locations at time n-1 and n; both point and interval predictions are provided

Usage

```
Predictions.COST.t(par,Y,s.ob,s.new,isotropic)
```

Arguments

par	parameters in the copula function
Y	observed data
s.ob	coordinates of observed locations
s.new	coordinates of new locations
isotropic	indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`Predictions.GP` *new location prediction by Gaussian process method*

Description

new location prediction by Gaussian process method, and the marginal mean and variance of the new location is estimated by neighboring information; it gives 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n, conditional on observed locations at time n-1 and n; both point and interval predictions are provided

Usage

```
Predictions.GP(par,Y,s.ob,s.new,isotropic)
```

Arguments

`par`	parameters in the copula function
`Y`	observed data
`s.ob`	coordinates of observed locations
`s.new`	coordinates of new locations
`isotropic`	indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`rank.multivariate` *multivariate rank of a vector*

Description

calculating the multivariate rank of a vector among a set of vectors, used to evaluate the performance of conditional distribution, and the rank would be uniform when the conditional distribution is estimated well

Usage

```
rank.multivariate(y.test,y.random,seed1)
```

Arguments

`y.test`	the observed (verifying) vector at time n+1
`y.random`	m random draws from the conditional distribution
`seed1`	random seed to solve tie at random

Value

the multivariate rank of the observed (verifying) vector at time n+1

Author(s)

Yanlin Tang and Huixia Judy Wang

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

`Wind6month` *Wind speed data from 10 sites*

Description

The data set is a subset of the data we used in the paper, with 10 sites and 6-month long time series.

Usage

```
data(Wind6month)
```

Format

A 4320*10 matrix from 10 locations, date ranges from Sep 22, 2014 to Dec 20, 2014, 180 days

`BiddleButte` wind speed from site BiddleButte

`ForestGrove` wind speed from site ForestGrove

`HoodRiver` wind speed from site HoodRiver

`HorseHeaven` wind speed from site HorseHeaven

`Megler` wind speed from site Megler

`NaselleRidge` wind speed from site NaselleRidge

`Roosevelt` wind speed from site Roosevelt

`Shaniko` wind speed from site Shaniko

`Sunnyside` wind speed from site Sunnyside

`Tillamook` wind speed from site Tillamook

Source

https://transmission.bpa.gov/business/operations/wind/MetData/default.aspx

References

Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based semiparametric models for spatio-temporal data.

Examples

```
data(Wind6month)
Y.ob = Wind6month[,-3]
Y.newloc = Wind6month[,3]
dim(Y.ob) #4320*9, data at 9 locations, with length 4320 (hours)
length(Y.newloc) #4320, time series at the new location
```

Index