Package ‘CopCTS’

July 23, 2025

Type Package

Title Copula-Based Semiparametric Analysis for Time Series Data with Detection Limits

Version 1.0.0

Author Fuyuan David Li

Maintainer Fuyuan David Li <LFY@gwmail.gwu.edu>

Description Semiparametric estimation for censored time series with lower detection limit. The latent response is a sequence of stationary process with Markov property of order one.
Estimation of copula parameter(COPC) and Conditional quantile estimation are included for five available copula functions.
Copula selection methods based on L2 distance from empirical copula function are also included.

Depends copula

Imports msm,copBasic,methods

License GPL

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2018-12-03 10:52:46 UTC

Contents

condQestCopC *Conditional Quantile Estimation*

Description

Given estiamted copula with copula parameter and specified marginal distribution, obtain the conditional qth quantile of Y_n+1 given Y1,...,Yn.

Usage

```
condQestCopC(tao,Yc,d,delta,copula,cop=NULL,theta=NULL,nIS=10000,
MARGIN=NULL,MARGIN.inv=NULL,...)
```

Arguments

`tao`	the desired quantile level, a numeric value between 0 and 1.
`Yc`	the Nx1 vector of observed responses that are subject to lower detection limit.
`d`	the lower detection limit.
`delta`	the Nx1 vector of censoring indicator with 1 indicating uncensored and 0 indicating left censored.
`copula`	the input copula object with copula parameter plugged in. If specified, `cop` and `theta` can be omitted.
`cop`	the choice of copula function. There are currently five available copula funcitons, including Clayton copula, Gaussian copula, Gumbel copula, Joe copula and Frank copula. Specify one from "Clayton","Gaussian","Gumbel","Joe" and "Frank".
`theta`	the copula parameter.
`nIS`	the size for sequential importance sampling. The default is 10000.
`MARGIN`	the marginal distribution of the latent time series.
`MARGIN.inv`	the inverse marginal distribution of the latent time series.
`...`	additional parameters for the marginal distribution of the latent time series.

Value

`condQestCopC` returns the conditional tao-th quantile of Y_n+1 given Y1,...,Yn based on the specified copula function and marginal distribution.

References

Li, F., Tang, Y. and Wang, H. (2018). Copula-Based Semiparametric Analysis for Time Series Data with Detection Limits, technical report.

Examples

```
set.seed(20)
Y = genLatentY(cop = "Clayton", theta = 1, N = 30)
d = -0.5
delta = (Y>d)
Yc = pmax(d,Y)
cq60.real = condQestCopC(0.6,Yc,d,delta,copula=claytonCopula(1),nIS = 50,
                        MARGIN=pnorm,MARGIN.inv=qnorm)
### Use selected copula
selCopC = selectCopC(cop.type = c("Clayton","Frank"),Yc,d,delta,nIS=50)
cq60.est = condQestCopC(0.6,Yc,d,delta,selCopC$Selected,nIS=50)
```

`estCopC` *Pseudo maximum likelihood estimator of the copula parameter*

Description

Obtains the pseudo maximum likelihood estimator of the copula parameter based on censored time series.

Usage

```
estCopC(cop="Gaussian",Yc,d,delta,nIS=500,jumps=NULL,MARGIN=NULL,...,interval=NULL)
```

Arguments

`cop`	the choice of copula function. There are currently five available copula funcitons, including Clayton copula, Gaussian copula, Gumbel copula, Joe copula and Frank copula. Specify one from "Clayton","Gaussian","Gumbel","Joe" and "Frank". The default is "Gaussian".
`Yc`	the Nx1 vector of observed response variable that is subject to lower detection limit.
`d`	the lower detection limit.
`delta`	the Nx1 vector of censoring indicator with 1 indicating uncensored and 0 indicating left censored.
`nIS`	the size for sequential importance sampling. The default is 500.
`jumps`	the Nx1 vector indicating whether each time t is a start of a new time series, which is deemed to be independent from the previous series. By default, `jumps` = c(1,rep(0,n-1)) indicating the data is one Markov sequence.
`MARGIN`	the marginal distribution function of the latent time series. The default is the empirical cdf: $$\frac{1}{n+1}\sum_{t=1}^{n} I_{Y_t<=y}$$. MARGIN can also be specified as other existing distribution functions such as pnorm.

...	additional parameters for the marginal distribution of the latent time series.
interval	the lower and upper bound for the copula paraameter. By default, `interval=` c(-1,1) for Gaussian copula, c(-1,Inf) for Clayton copula, c(1,Inf) for Gumbel and Joe copula and c(-Inf,Inf) for Frank copula.

Value

`estCopC` returns a list of components including.

para	the pseudo maximum likelihood estimator of the copula parameter.
likelihood	the negative log-likelihood value corresponding to the estimated copula parameter.
copula	the estimated copula object, with estimated copula parameter plugged in.

References

Li, F., Tang, Y. and Wang, H. (2018). Copula-based Semiparametric Analysis for Time Series Data with Detection Limits, technical report.

Examples

```
### Using a simulated data for demonstration:
set.seed(20)
Y = genLatentY(cop="Clayton",1,30,MARGIN.inv = qt,df=3)
d = -1
Yc = pmax(d,Y)
delta = (Y>d)
## CopC estimator
estCopC(cop = "Clayton",Yc,d,delta,nIS = 50,interval = c(1,10))
## Omniscient estimator
estCopC(cop = "Clayton",Y,d,delta=rep(TRUE,length(Y)),interval = c(1,10))
## CopC estimator under true marginal
estCopC(cop = "Clayton",Yc,d,delta,nIS = 50,MARGIN=pt,df=3,interval = c(1,10))
### Analyze the water quality data:
attach(water)
Yc = TNH3[1:30]
delta = Delta[1:30]
jumps = Indep[1:30]
set.seed(1)
estCopC(cop="Clayton",Yc=Yc,d=0.02,delta=delta,jumps=jumps,interval = c(1,10),nIS=50)
```

---

genLatentY *Generation of data from the copula-based Markov model of order one*

Description

Generate the latent response variable from the assumed copula-based Markov model in Li, Tang and Wang (2018).

Usage

```
genLatentY(cop="Gaussian",theta,N,MARGIN.inv=qnorm,...)
```

Arguments

cop	the choice of copula function. There are currently five available copula funci-tons, including Clayton copula, Gaussian copula, Gumbel copula, Joe copula and Frank copula. Specify one from "Clayton","Gaussian","Gumbel","Joe" and "Frank". The default is "Gaussian".
theta	the copula parameter.
N	the length of the latent response.
MARGIN.inv	the inverse marginal distribution function of the latent time series. The default is qnorm(p,mean=0,sd=1), i.e., the standard normal marginal.
...	additional parameters for the inverse marginal distribution funcion of the latent time series.

Value

genLatentY returns a Nx1 vector of the latent response variable Y*

References

Li, F., Tang, Y. and Wang, H. (2018) Copula-based Semiparametric Analysis for Time Series Data with Detection Limits, technical report.

---

selectCopC *The selection of copula function*

Description

Among a list of copulas, select the one that gives the estimates closest to the empirical copula function.

Usage

```
selectCopC(cop.type=c("Clayton","Gaussian","Gumbel","Joe","Frank"),
Yc,d,delta,nIS=500,jumps=NULL,MARGIN=NULL,...,intervals=NULL)
```

Arguments

cop.type	a Kx1 vector containing the candidate copulas, where K = length(cop.type) is the number of candidate copulas. There are currently five available copula funci-tons, including Clayton copula, Gaussian copula, Gumbel copula, Joe copula and Frank copula. Select each by specifying a vector consisting of at least one element from c("Clayton","Gaussian","Gumbel","Joe","Frank").
Yc	the Nx1 vector of observed responses that are subject to lower detection limit.

`d`	the lower detection limit.
`delta`	the Nx1 vector of censoring indicator with 1 indicating uncensored and 0 indicating left censored.
`nIS`	the size for sequential importance sampling. The default is 500.
`jumps`	the Nx1 vector indicating whether each time t is a start of a new time series, which is deemed to be independent from the previous series.
`MARGIN`	the marginal distribution of the latent time series.
`...`	additional parameters for the marginal distribution of the latent time series.
`intervals`	a 2xK matrix specifying the lower and upper bound for the copula parameter of each candidate copula, where K is the number of candidate copulas.

Value

`selectCopC` returns a list of components including

`paras`	a Kx1 vector containing the estimated copula parameters for each candidate copula.
`likelihoods`	a Kx1 vector containing the negative log-likelihood value corresponding to the estimated copula parameter for each candidate copula.
`estCop`	a list containing the estimated copula object for each candidate.
`L2distance`	a Kx1 vector containing the L2 distance between each copula with estimated copula parameter and the empirical copula function.
`Selected`	The selected copula object.

References

Li, F., Tang, Y. and Wang, H. (2018) Copula-based Semiparametric Analysis for Time Series Data with Detection Limits, technical report.

See Also

`estCopC`.

Examples

```
### Example with simulated data
set.seed(20)
Y = genLatentY("Clayton",1,30,MARGIN.inv = qt,df=3)
d = -1
Yc = pmax(d,Y)
delta = (Y>d)
selectCopC(cop.type=c("Clayton","Frank"),Yc = Yc,d = d,delta = delta,nIS=50)
### Example with water data
attach(water)
Yc = TNH3[1:30]
delta = Delta[1:30]
jumps = Indep[1:30]
set.seed(1)
```

```
intv.Gaussian = c(-1,1)
intv.Clayton = c(0,20)
intv.Frank = c(0,15)
intervals = cbind(intv.Gaussian,intv.Clayton,intv.Frank)
cop.type = c("Gaussian","Clayton","Frank")
selCopC <- selectCopC(cop.type=cop.type,Yc=Yc,d=0.02,
            delta=delta,nIS = 50,jumps=jumps,intervals=intervals)
selCopC$Selected
```

`water` *Water quality (Ammonia) data*

Description

This water dataset records the amount of dissolved ammonia at Susquehanna River Basin in the United States. The dissolved ammonia data were observed biweekly in Susquehanna River at Towanda, PA, from 1988 to 2014, consisting of 524 data points, with detection limit at 0.02 (mg/l).

Usage

```
data(water)
```

Format

A data frame with 524 observations on the following 4 variables.

- `SDate` date of measuring
- `TNH3` response variable, the amount of dissolved ammonia
- `Delta` a logical vector indicating censored as 0 and uncensored as 1
- `Indep` a logical vector indicating the start of a new time series that is deemed to be independent from the previous one. For the water quality data, most measurements were taken biweekly but a few have longer time gaps from the previous measurements. In our analysis of the water quality data, we treat the date that is apart from the previous measurement date more than 14 days as the start of a new independent time series.

Source

https://www.srbc.net/portals/water-quality-projects/sediment-nutrient-assessment/

References

Li, F., Tang, Y. and Wang, H. (2018) Copula-based Semiparametric Analysis for Time Series Data with Detection Limits, technical report.

Examples

```
data(water)
str(water)
head(water)
```

Index