Package 'BRBVS'

January 20, 2025

Title Variable Selection and Ranking in Copula Survival Models
Affected by General Censoring Scheme

Version 0.2.1

Description Performs variable selection and ranking based on several measures for the class of copula survival model(s) in high dimensional domain.
The package is based on the class of copula survival model(s) implemented in the 'GJRM' package.

License GPL (>= 3)

Depends R (>= 3.5)

Imports copent (>= 0.4), ggplot2 (>= 3.4.3), GJRM (>= 0.2-6.4),
mvtnorm (>= 1.2-4), stats, utils

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

NeedsCompilation no

Author Danilo Petti [aut, cre, cph] (<https://orcid.org/0000-0003-1913-6891>),
Marcella Niglio [aut] (<https://orcid.org/0000-0002-4220-4277>),
Marialuisa Restaino [aut]

Maintainer Danilo Petti <d.petti@essex.ac.uk>

Repository CRAN

Date/Publication 2024-06-27 16:50:02 UTC

Contents

AREDS   *Age-Related Eye Disease Study (AREDS)*

Description

A real dataset of bivariate interval-censored data with 629 subjects and 4 non-genetic covariates and 1 genetic covariate. The dataset is selected from the Age-related Eye Disease Study (AREDS) (AREDS Group, 1999). Two events are the progression times (in years) to late-AMD in the left and right eyes.

Usage

```
AREDS
```

Format

AREDS:

A data frame with 628 patients (rows) and 11 columns:

SevScale1E Severity scale associated with the right eye

SevScale2E Severity scale associated with the left eye

ENROLLAGE Age at baseline

rs2284665 SNP covariate highly associated with late-AMD progression (GG, GT, TT)

cens1 Recurrence censoring variable in the right eye

cens2 Recurrence censoring variable in the left eye

cens Overall survival censoring variable

t11 Start of follow-up in days right eye

t12 Time to recurrence or last follow-up in days right eye

t21 Start of follow-up in days left eye

t22 Time to recurrence or last follow-up in days left eye ...

Source

https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v3.p1

---

backward_selection_BivCop

Backward Selection for Bivariate Copula Survival Models

Description

This function performs backward selection based on AIC or BIC measures for bivariate copula survival models. It iteratively removes variables from the model to minimize the specified measure, either AIC or BIC.

Usage

```
backward_selection_BivCop(
  data,
  lowerBt1 = "t11",
  lowerBt2 = "t21",
  upperBt1 = "t12",
  upperBt2 = "t22",
  copula = "N",
  margins = c("PH", "PH"),
  measure = "AIC",
  cens1,
  cens2
)
```

Arguments

data	A data frame containing the dataset.
lowerBt1	Character. Name of the lower bound for the first time to event.
lowerBt2	Character. Name of the lower bound for the second time to event.
upperBt1	Character. Name of the upper bound for the first time to event.
upperBt2	Character. Name of the upper bound for the second time to event.
copula	Character. Type of copula to be used in the model. Default is 'N' (Normal copula).
margins	Character vector. Margins to be used in the copula model. Default is c('PH', 'PH').
measure	Character. Measure to be minimized during the selection process. Either 'AIC' or 'BIC'. Default is 'AIC'.
cens1	Censoring indicator for the first time to event.
cens2	Censoring indicator for the second time to event.

Value

A list containing:

- `Results`: A data frame with the steps, models, and the corresponding AIC/BIC values.
- `Equations`: A list with the final model equations for the selected variables.

Examples

```
#################################################
# Example based on AREDS dataset
# This analysis serves solely as a
# demonstration of the function's capabilities.
#################################################
data(AREDS)
subsetAREDS <- AREDS[, c('t11', 't12', 't21', 't22', 'SevScale1E',
                         'SevScale2E', 'cens1', 'cens2', 'cens')]
results <- backward_selection_BivCop(data = subsetAREDS, lowerBt1 = 't11', lowerBt2 = 't21',
                                     upperBt1 = 't12', upperBt2 = 't22',
                                     copula = 'N', margins = c('PH', 'PH'),
                                     measure = 'AIC', cens1 = AREDS$cens1,
                                     cens2 = AREDS$cens2)
print(results)
```

BRBVS   *Bivariate Rank-Based Variable Selection*

Description

This function performs bivariate rank-based variable selection (BRBVS) based on copula survival copula models. It computes rankings for covariates and selects a specified number of variables according to the estimated probabilities. The function returns rankings and selected variables for different criteria.

Usage

```
BRBVS(y, x, kmax, copula, margins, m, tau, n.rep, metric)
```

Arguments

y	Time to events and censoring matrix as a data frame.
x	Covariates matrix as a data frame. Input matrix containing the predictor variables.
kmax	Numeric. The maximum number of variables to be selected. Must be positive, non-zero, and less than or equal to the number of columns in x.
copula	Character. Type of copula employed in the algorithm. Must be one of the following types: N, C0, C90, C180, C270, GAL0, GAL90, GAL180, GAL270, J0, J90, J180, J270, G0, G90, G180, G270, F, AMH, FGM, T, PL, HO default is C0. See GJRM package documentation for details: GJRM package.
margins	Character. Type of margin employed in the algorithm. Must be one of PH, PO, probit. Default is c(PH, PO). See GJRM package documentation for more on margins: GJRM package.

m	Numeric. Subsample size, typically set to n/2 where n is the number of observations.
tau	Numeric. A user-defined threshold for variable selection. Must be in the interval (0,1), exclusive.
n.rep	Integer. Number of Bootstrap replicates. Must be positive.
metric	Character, specifies the metric used for ranking the variables. Must be one of 'CE', 'FIM', 'Abs'. Default is 'FIM'.

Value

A list containing the following components:

- `mtx.act1E`: Numeric vector of indices of the active variables selected based on the first survival function. Remaining positions (up to 'kmax') are filled with 0.
- `score.r1E`: Numeric vector of the ranked scores for variable selection based on the first survival function, with remaining positions (up to 'kmax'-1) filled with 0.
- `freq.rel1E`: Numeric vector of the relative frequencies of selected variables based on the first survival function (frequencies divided by 'n.rep').
- `mtx.act2E`: Numeric vector of indices of the active variables selected based on the second survival function. Remaining positions (up to 'kmax') are filled with 0.
- `score.r2E`: Numeric vector of the ranked scores for variable selection based on the second survival function, with remaining positions (up to 'kmax'-1) filled with 0.
- `freq.rel2E`: Numeric vector of the relative frequencies of selected variables based on the second survival function (frequencies divided by 'n.rep').
- `metric`: The metric used for ranking the variables.
- `kmax`: The maximum number of variables to be selected.
- `copula`: The type of copula employed in the algorithm.
- `margins`: The type of margins employed in the algorithm. . - Namecondings: Table with name of covariates and encoding used in the output.

Examples

```
################################################
# Example based on AREDS dataset
# This analysis serves solely as a
# demonstration of the function's capabilities.
################################################

data(AREDS)
Y<- AREDS[,c('t11','t12', 't21', 't22', 'cens1', 'cens2', 'cens')]
X<- AREDS[,c(3, 9)]
# Including just 1 covariates as example
X$SevScale1E <- scale(as.numeric( X$SevScale1E))
X$SevScale2E <- scale(as.numeric(X$SevScale1E))


Bivrbvs<- BRBVS(y=Y, x=X, kmax=2,copula='C0',
```

```
                              margins=c('PO','PO'),
                              m=628 , # try to set m=628 (628 is the sample size)
                              tau=0.5,
                              n.rep=1, # number of bootstrap = 1
                              metric='FIM')
```

---

forward_selection_BivCop

Forward Selection for Bivariate Copula Survival Models

Description

This function performs forward selection based on AIC or BIC measures for bivariate copula survival models. It iteratively adds variables to the model to minimize the specified measure, either AIC or BIC.

Usage

```
forward_selection_BivCop(
  data,
  lowerBt1 = "t11",
  lowerBt2 = "t21",
  upperBt1 = "t12",
  upperBt2 = "t22",
  copula = "N",
  margins = c("PH", "PH"),
  measure = "AIC",
  cens1,
  cens2
)
```

Arguments

data	A data frame containing the dataset.
lowerBt1	Character. Name of the lower bound for the first time to event.
lowerBt2	Character. Name of the lower bound for the second time to event.
upperBt1	Character. Name of the upper bound for the first time to event.
upperBt2	Character. Name of the upper bound for the second time to event.
copula	Character. Type of copula to be used in the model. Default is 'N' (Normal copula).

margins	Character vector. Margins to be used in the copula model. Default is c('PH', 'PH').
measure	Character. Measure to be minimized during the selection process. Either 'AIC' or 'BIC'. Default is 'AIC'.
cens1	Censoring indicator for the first time to event.
cens2	Censoring indicator for the second time to event.

Value

A list containing:

- `Results`: A data frame with the steps, models, and the corresponding AIC/BIC values.

- `Equations`: A list with the final model equations for the selected variables.

Examples

```
################################################
# Example based on AREDS dataset
# This analysis serves solely as a
# demonstration of the function's capabilities.
################################################
data(AREDS)
subsetAREDS <- AREDS[, c('t11', 't12', 't21', 't22', 'SevScale1E',
                         'SevScale2E', 'cens1', 'cens2', 'cens')]
results <- forward_selection_BivCop(data = subsetAREDS, lowerBt1 = 't11', lowerBt2 = 't21',
                                    upperBt1 = 't12', upperBt2 = 't22',
                                    copula = 'N', margins = c('PH', 'PH'),
                                    measure = 'AIC', cens1 = AREDS$cens1,
                                    cens2 = AREDS$cens2)
print(results)
```

---

plotBRBVS *Plot Results of BRBVS Algorithm*

Description

This function takes an object containing the results of the BRBVS algorithm and plots the active features against their relative frequencies for the two time to events.

Usage

```
plotBRBVS(object, ...)
```

Arguments

object   An object containing the results of the BRBVS algorithm. Expected to contain
                matrices `mtx.act1E` and `mtx.act2E` for feature names, vectors `freq.rel1E` and
                `freq.rel2E` for relative frequencies, and additional elements `Namecondings`,
                `metric`, `copula`, `margins` for feature decoding and plot annotations.

...   Additional arguments to be passed to plotting functions.

Value

A ggplot object representing the BRBVS algorithm results. Features are displayed on the x-axis
and their relative frequencies on the y-axis. The plot is split into two facets for each survival.

Examples

```
#############################
# Toy Example
#############################
BRBVS_object <- list(
mtx.act1E = c("z2", "z1", "0", "0"),
scores1E = c(20, 1, 1),
freq.rel1E = c(1, 1, 1, 1),
mtx.act2E = c("z2", "z1", "0", "0"),
scores2E = c(20, 1, 1),
freq.rel2E = c(1, 1, 1, 1),
metric = "FIM",
kmax = 4,
copula = "C0",
margins = c("PO", "PO"),
tau = 0.5,
Namecondings = matrix(
 c("z1", "z1", "z2", "z2", "z3", "z3", "z4", "z4", "z5", "z5", "z6", "z6"),
 nrow = 6, ncol = 2, byrow = TRUE,
 dimnames = list(NULL, c("Features", "Code"))
)
)

class(BRBVS_object) <- "BRBVS"
plotBRBVS(BRBVS_object)
```

Select_link_BivCop   *Select Best Link Function for Bivariate Copula Survival Models*

Description

Selects the best link function for bivariate copula survival models based on AIC or BIC measures.
It evaluates different margins (link functions) for the survival models and selects the one with the
lowest AIC or BIC.

Usage

```
Select_link_BivCop(
  data,
  cens1,
  cens2,
  lowerBt1 = "t11",
  lowerBt2 = "t21",
  upperBt1 = "t12",
  upperBt2 = "t22",
  measure = "AIC",
  eta1 = NULL,
  eta2 = NULL,
  input_equation = FALSE
)
```

Arguments

data	A data frame containing the dataset.
cens1	Censoring indicator for the first time to event.
cens2	Censoring indicator for the second time to event.
lowerBt1	Character. Name of the lower bound for the first time to event.
lowerBt2	Character. Name of the lower bound for the second time to event.
upperBt1	Character. Name of the upper bound for the first time to event.
upperBt2	Character. Name of the upper bound for the second time to event.
measure	Character. Measure to be minimized during the selection process. Either 'AIC' or 'BIC'. Default is 'AIC'.
eta1	Formula for the first survival model equation. Default is NULL.
eta2	Formula for the second survival model equation. Default is NULL.
input_equation	Logical. If TRUE, uses the provided eta1 and eta2 formulas. If FALSE, generates formulas using all predictors in data. Default is FALSE.

Value

A list containing:

- best_margin_S1: The best margin (link function) for the first survival model.

- measure: The AIC or BIC value for the best margin of the first survival model.

- model_S1: The fitted model for the first survival model with the best margin.

- best_margin_S2: The best margin (link function) for the second survival model.

- measure_S2: The AIC or BIC value for the best margin of the second survival model.

- model_S2: The fitted model for the second survival model with the best margin.

Examples

```
#################################################
# Example based on AREDS dataset
# This analysis serves solely as a
# demonstration of the function's capabilities.
#################################################
data(AREDS)

results <- Select_link_BivCop(data = AREDS, cens1 = AREDS$cens1,
                             lowerBt1 = 't11', lowerBt2 = 't21',
                             upperBt1 = 't12', upperBt2 = 't22',
                             cens2 = AREDS$cens2, measure = 'AIC')
print(results)
```

---

summary.BRBVS   *Summary for BRBVS Object*

Description

Provides a concise overview of a Bivariate Variable Selection (BRBVS) object. It reports the maximum number of relevant covariates (kmax) across all sets, and details the relevant covariates for each survival function along with their frequency of selection.

Usage

```
## S3 method for class 'BRBVS'
summary(object, ...)
```

Arguments

object	An object of class BRBVS, typically the result of a BRBVS analysis.
...	Further arguments passed to or from other methods.

Value

This function does not return a value but prints the summary of the BRBVS object to the console. If the BRBVS object is invalid or incomplete, an error message is displayed and the function returns NULL.

Examples

```
#############################
# Toy Example
#############################
BRBVS_object <- list(
mtx.act1E = c("z2", "z1", "0", "0"),
```

```
scores1E = c(20, 1, 1),
freq.rel1E = c(1, 1, 1, 1),
mtx.act2E = c("z2", "z1", "0", "0"),
scores2E = c(20, 1, 1),
freq.rel2E = c(1, 1, 1, 1),
metric = "FIM",
kmax = 4,
copula = "C0",
margins = c("PO", "PO"),
tau = 0.5,
Namecondings = matrix(
 c("z1", "z1", "z2", "z2", "z3", "z3", "z4", "z4", "z5", "z5", "z6", "z6"),
 nrow = 6, ncol = 2, byrow = TRUE,
 dimnames = list(NULL, c("Features", "Code"))
)
)

class(BRBVS_object) <- "BRBVS"
summary(BRBVS_object)
```

Index