Package ‘FactorCopula’

September 11, 2020

Version 0.5

Date 2020-08-30

Title Factor Copula Models for Mixed Continuous and Discrete Data

Author Sayed H. Kadhem [aut, cre],
Aristidis K. Nikoloulopoulos [aut]

Maintainer Sayed H. Kadhem <s.kadhem@uea.ac.uk>

Depends R (>= 3.5.0), statmod, abind, utils, polycor

Description Estimation, model selection and goodness-of-fit of factor copula models for mixed continuous and discrete data in Kadhem and Nikoloulopoulos (2019) <arXiv:1907.07395>.

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-09-11 09:40:11 UTC

R topics documented:

FactorCopula-package *Factor copula models for mixed continuous and disrete data*

Description

Estimation, model selection and goodness-of-fit of factor copula models for mixed continuous and discrete data in Kadhem and Nikoloulopoulos (2019) <arXiv:1907.07395>.

Details

This package contains R functions for:

- diagnostics based on semi-correlations (Kadhem and Nikoloulopoulos, 2019; Joe, 2014) to detect tail dependence or tail asymmetry;
- diagnostics to show that a dataset has a factor structure based on linear factor analysis (Kadhem and Nikoloulopoulos, 2019; Joe, 2014);
- estimation of the factor copula models in Krupskii and Joe (2013), Nikoloulopoulos and Joe (2015), and Kadhem and Nikoloulopoulos (2019);
- model selection of the factor copula models in Krupskii and Joe (2013), Nikoloulopoulos and Joe (2015), and Kadhem and Nikoloulopoulos (2019) using the heuristic algorithm in Kadhem and Nikoloulopoulos (2019) that automatically selects the bivariate parametric copula families that link the observed to the latent variables;
- goodness-of-fit of the factor copula models in Krupskii and Joe (2013), Nikoloulopoulos and Joe (2015), and Kadhem and Nikoloulopoulos (2019) using the M_2 statistic (Maydeu-Olivares and Joe, 2006). Note that the continuous and count data have to be transformed to ordinal.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References

Joe, H. (2014). *Dependence Modelling with Copulas*. Chapman and Hall/CRC.

Maydeu-Olivares, A. and Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, **71**, 713–732. doi: 10.1007/s1133600512959.

Kadhem, S.H. and Nikoloulopoulos, A.K. (2019) Factor copula models for mixed data. *Arxiv e-prints*, <arXiv:1907.07395>. https://arxiv.org/abs/1907.07395.

Krupskii, P. and Joe, H. (2013) Factor copula models for multivariate data. *Journal of Multivariate Analysis*, **120**, 85–101. doi: 10.1016/j.jmva.2013.05.001.

Nikoloulopoulos, A.K. and Joe, H. (2015) Factor copula models with item response data. *Psychometrika*, **80**, 126–150. doi: 10.1007/s1133601393874.

discrepancy *Diagnostics to detect a factor dependence structure.*

Description

The diagnostic method in Joe (2014, pages 245-246) to show that each dataset has a factor structure based on linear factor analysis. The correlation matrix $\mathsf{R}_{\text{observed}}$ has been obtained based on the sample correlations from the bivariate pairs of the observed variables. These are the linear (when both variables are continuous), polychoric (when both variables are ordinal), and polyserial (when one variable is continuous and the other is ordinal) sample correlations among the observed variables. The resulting $\mathsf{R}_{\text{observed}}$ is generally positive definite if the sample size is not small enough; if not one has to convert it to positive definite. We calculate various measures of discrepancy between $\mathsf{R}_{\text{observed}}$ and $\mathsf{R}_{\text{model}}$ (the resulting correlation matrix of linear factor analysis), such as the maximum absolute correlation difference $D_1 = \max|\mathsf{R}_{\text{model}} - \mathsf{R}_{\text{observed}}|$, the average absolute correlation difference $D_2 = \text{avg}|\mathsf{R}_{\text{model}} - \mathsf{R}_{\text{observed}}|$, and the correlation matrix discrepancy measure $D_3 = \log\big(\det(\mathsf{R}_{\text{model}})\big) - \log\big(\det(\mathsf{R}_{\text{observed}})\big) + \text{tr}(\mathsf{R}_{\text{model}}^{-1}\mathsf{R}_{\text{observed}}) - d$.

Usage

```
discrepancy(cormat, n, f3)
```

Arguments

cormat	$\mathsf{R}_{\text{observed}}$.
n	Sample size.
f3	If TRUE, then the linear 3-factor analysis is fitted.

Value

A matrix with the calculated discrepancy measures for different number of factors.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References

Joe, H. (2014). *Dependence Modelling with Copulas*. Chapman and Hall/CRC.

Kadhem, S.H. and Nikoloulopoulos, A.K. (2019) Factor copula models for mixed data. *Arxiv e-prints*, <arXiv:1907.07395>. https://arxiv.org/abs/1907.07395.

Examples

```
#-----------------------------------------------
#                    PE Data
#------------------            -----------------
data(PE)
#correlation
continuous.PE1 <- -PE[,1]
continuous.PE <- cbind(continuous.PE1, PE[,2])
u.PE <- apply(continuous.PE, 2, rank)/(nrow(PE)+1)
z.PE <- qnorm(u.PE)
categorical.PE <- data.frame(apply(PE[, 3:5], 2, factor))
nPE <- cbind(z.PE, categorical.PE)

#------------------------------------------------
# Discrepancy measures----------------------------
#------------------------------------------------
#correlation matrix for mixed data
cormat.PE <- as.matrix(polycor::hetcor(nPE, std.err=FALSE))
#discrepancy measures
out.PE = discrepancy(cormat.PE, n = nrow(nPE), f3 = FALSE)

#-----------------------------------------------
#-----------------------------------------------
#                   GSS Data
#------------------            -----------------
data(GSS)
attach(GSS)
continuous.GSS <- cbind(INCOME,AGE)
continuous.GSS <- apply(continuous.GSS, 2, rank)/(nrow(GSS)+1)
z.GSS <- qnorm(continuous.GSS)
ordinal.GSS <- cbind(DEGREE,PINCOME,PDEGREE)
count.GSS <- cbind(CHILDREN,PCHILDREN)

# Transforming the count variables to ordinal
# count1 : CHILDREN
count1  = count.GSS[,1]
count1[count1 > 3] = 3

# count2: PCHILDREN
count2  = count.GSS[,2]
count2[count2 > 7] = 7

# Combining both transformed count variables
ncount.GSS = cbind(count1, count2)

# Combining ordinal and transformed count variables
categorical.GSS <- cbind(ordinal.GSS, ncount.GSS)
categorical.GSS <- data.frame(apply(categorical.GSS, 2, factor))

# combining continuous and categorical variables
nGSS = cbind(z.GSS, categorical.GSS)
```

```
#-----------------------------------------------
# Discrepancy measures----------------------------
#-----------------------------------------------
#correlation matrix for mixed data
cormat.GSS <- as.matrix(polycor::hetcor(nGSS, std.err=FALSE))
#discrepancy measures
out.GSS = discrepancy(cormat.GSS, n = nrow(nGSS), f3 = TRUE)
```

GSS *The 1994 General Social Survey*

Description

Hoff (2007) analysed seven demographic variables of 464 male respondents to the 1994 General Social Survey. Of these seven, two were continuous (income and age of the respondents), three were ordinal with 5 categories (highest degree of the survey respondent, income and highest degree of respondent's parents), and two were count variables (number of children of the survey respondent and respondent's parents).

Usage

```
data(GSS)
```

Format

A data frame with 464 observations on the following 7 variables:

`INCOME` Income of the respondent in 1000s of dollars, binned into 21 ordered categories.

`DEGREE` Highest degree ever obtained (0:None, 1:HS, 2:Associates, 3:Bachelors, 4:Graduate).

`CHILDREN` Number of children of the survey respondent.

`PINCOME` Financial status of respondent's parents when respondent was 16 (on a 5-point scale).

`PDEGREE` Highest degree of the survey respondent's parents (0:None, 1:HS, 2:Associates, 3:Bachelors, 4:Graduate).

`PCHILDREN` Number of children of the survey respondent's parents - 1.

`AGE` Age of the respondents in years.

Source

Hoff, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *The Annals of Applied Statistics*, **1**, 265–283.

M2 *Goodness-of-fit of factor copula models for mixed data*

Description

The limited information M_2 statistic (Maydeu-Olivares and Joe, 2006) of factor copula models for mixed continuous and discrete data.

Usage

```
M2.1F(tcontinuous, ordinal, tcount, cpar, copF1, gl)
M2.2F(tcontinuous, ordinal, tcount, cpar, copF1, copF2, gl, SpC)
```

Arguments

`tcontinuous`	$n \times d_1$ matrix with the transformed continuous to ordinal reponse data, where n and d_1 is the number of observations and transformed continous variables, respectively.
`ordinal`	$n \times d_2$ matrix with the ordinal reponse data, where n and d_2 is the number of observations and ordinal variables, respectively.
`tcount`	$n \times d_3$ matrix with the transformed count to ordinal reponse data, where n and d_3 is the number of observations and transformed count variables, respectively.
`cpar`	A list of estimated copula parameters.
`copF1`	$(d_1 + d_2 + d_3)$-vector with the names of bivariate copulas that link the each of the oberved variabels with the 1st factor. Choices are "bvn" for BVN, "bvtν" with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, "frk" for Frank, "gum" for Gumbel, "rgum" for reflected Gumbel, "1rgum" for 1-reflected Gumbel, "2rgum" for 2-reflected Gumbel, "joe" for Joe, "rjoe" for reflected Joe, "1rjoe" for 1-reflected Joe, "2rjoe" for 2-reflected Joe, "BB1" for BB1, "rBB1" for reflected BB1, "BB7" for BB7, "rBB7" for reflected BB7,"BB8" for BB8, "rBB8" for reflected BB8, "BB10" for BB10, "rBB10" for reflected BB10.
`copF2`	$(d_1 + d_2 + d_3)$-vector with the names of bivariate copulas that link the each of the oberved variabels with the 2nd factor. Choices are "bvn" for BVN, "bvtν" with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, "frk" for Frank, "gum" for Gumbel, "rgum" for reflected Gumbel, "1rgum" for 1-reflected Gumbel, "2rgum" for 2-reflected Gumbel, "joe" for Joe, "rjoe" for reflected Joe, "1rjoe" for 1-reflected Joe, "2rjoe" for 2-reflected Joe, "BB1" for BB1, "rBB1" for reflected BB1, "BB7" for BB7, "rBB7" for reflected BB7,"BB8" for BB8, "rBB8" for reflected BB8, "BB10" for BB10, "rBB10" for reflected BB10.
`gl`	Gauss legendre quardrature nodes and weights.
`SpC`	Special case for the 2-factor copula model with BVN copulas. Select a bivariate copula at the 2nd factor to be fixed to independence. e.g. "SpC = 1" to set the first copula at the 2nd factor to independence.

Details

The M_2 statistic has been developed for goodness-of-fit testing in multidimensional contingency tables by Maydeu-Olivares and Joe (2006). Nikoloulopoulos and Joe (2015) have used the M_2 statistic to assess the goodness-of-fit of factor copula models for ordinal data. We build on the aforementioned papers and propose a methodology to assess the overall goodness-of-fit of factor copula models for mixed continuous and discrete responses. Since the M_2 statistic has been developed for multivariate ordinal data, we propose to first transform the continuous and count variables to ordinal and then calculate the M_2 statistic at the maximum likelihood estimate before transformation.

Value

A list containing the following components:

`M2` The M_2 statistic which has a null asymptotic distribution that is χ^2 with $s-q$ degrees of freedom, where s is the number of univariate and bivariate margins that do not include the category 0 and q is the number of model parameters.

`df` $s-q$.

`p-value` The resultant p-value.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References

Kadhem, S.H. and Nikoloulopoulos, A.K. (2019) Factor copula models for mixed data. *Arxiv e-prints*, <arXiv:1907.07395>. https://arxiv.org/abs/1907.07395.

Maydeu-Olivares, A. and Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, **71**, 713–732. doi: 10.1007/s1133600512959.

Nikoloulopoulos, A.K. and Joe, H. (2015) Factor copula models with item response data. *Psychometrika*, **80**, 126–150. doi: 10.1007/s1133601393874.

Examples

```
#------------------------------------------------
# Setting quadreture points
nq <- 25
gl <- gauss.quad.prob(nq)
#------------------------------------------------
#                    PE Data
#------------------            -----------------
data(PE)
continuous.PE1 = -PE[,1]
continuous.PE2 = PE[,2]
continuous.PE <- cbind(continuous.PE1, continuous.PE2)
```

```
categorical.PE <- PE[, 3:5]
#------------------------------------------------
#                   Estimation
#------------------             -----------------
#------------------ One-factor  -----------------
# one-factor copula model
cop1f.PE <- c("joe", "joe", "rjoe", "joe", "gum")
est1factor.PE <- mle1factor(continuous.PE, categorical.PE,
                            count=NULL, copF1=cop1f.PE, gl, hessian = T)
#------------------------------------------------
#                       M2
#------------------------------------------------
#Transforming the continuous to ordinal data:
ncontinuous.PE = continuous2ordinal(continuous.PE, 5)
# M2 statistic for the one-factor copula model:

m2.1f.PE <- M2.1F(ncontinuous.PE, categorical.PE, tcount=NULL,
       cpar=est1factor.PE$cpar, copF1=cop1f.PE, gl)

#------------------------------------------------
#                    GSS Data
#------------------             -----------------
data(GSS)
attach(GSS)
continuous.GSS <- cbind(INCOME,AGE)
ordinal.GSS <- cbind(DEGREE,PINCOME,PDEGREE)
count.GSS <- cbind(CHILDREN,PCHILDREN)

#------------------------------------------------
#                   Estimation
#------------------             -----------------
# one-factor copula model
cop1f.GSS <- c("joe","2rjoe","bvt3","bvt3",
          "rgum","2rjoe","2rgum")
est1factor.GSS <- mle1factor(continuous.GSS, ordinal.GSS,
                        count.GSS, copF1=cop1f.GSS, gl, hessian = T)

#two-factor copula model
cop1.2f <- c("rgum","rjoe","bvn","1rjoe",
           "1rjoe","rjoe","gum")
cop2.2f <- c("gum","2rjoe","rjoe","gum",
           "bvt5","bvn","2rgum")
est2factor.GSS <- mle2factor(continuous.GSS, ordinal.GSS,
        count.GSS, copF1=cop1.2f, copF2=cop2.2f, gl, hessian = T)

#------------------------------------------------
#                 Transformation
#------------------             -----------------
# Transforming the continuous to ordinal data:

# continuous1: Income
continuous1 = as.integer(cut(continuous.GSS[,1],
c(0,10,19,29,40,100), include.lowest = T))
```

```
continuous1 = continuous1 - 1

# continuous2: AGE
continuous2 = as.integer(cut(continuous.GSS[,2] ,
c(0, 24, 44, 64, 100), include.lowest = T))
continuous2 = continuous2 - 1

# Combining the transformed continuous variables.
ncontinuous.GSS <- cbind(continuous1, continuous2)

#---------------- COUNT VARIABLE -----------------
# count1 : CHILDREN
count1  = count.GSS[,1]
count1[count1 > 3] = 3

# count2: PCHILDREN
count2  = count.GSS[,2]
count2[count2 > 7] = 7

# Combining both transformed count variables
ncount.GSS = cbind(count1, count2)

#------------------------------------------------
#                       M2
#------------------------------------------------
# M2 statistic for the one-factor copula model:

m2.1f.GSS <- M2.1F(ncontinuous.GSS, ordinal.GSS,
                      ncount.GSS, cpar = est1factor.GSS$cpar,
                      copF1 = cop1f.GSS, gl)

#------------------------------------------------
# M2 statistic for the two-factor copula model:

m2.2f.GSS <- M2.2F(ncontinuous.GSS, ordinal.GSS, ncount.GSS,
                      cpar = est2factor.GSS$cpar, copF1 = cop1.2f,
                      copF2 = cop2.2f, gl)
```

`mapping` *Mapping of Kendall's tau and copula parameter*

Description

Bivariate copulas: mapping of Kendall's tau and copula parameter.

Usage

```
par2tau(copulaname, cpar)
tau2par(copulaname, tau)
```

Arguments

copulaname
: Choices are "bvn" for BVN, "bvtν" with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, "frk" for Frank, "gum" for Gumbel, "rgum" for reflected Gumbel, "1rgum" for 1-reflected Gumbel, "2rgum" for 2-reflected Gumbel, "joe" for Joe, "rjoe" for reflected Joe, "1rjoe" for 1-reflected Joe, "2rjoe" for 2-reflected Joe, "BB1" for BB1, "rBB1" for reflected BB1, "BB7" for BB7, "rBB7" for reflected BB7, "BB8" for BB8, "rBB8" for reflected BB8, "BB10" for BB10, "rBB10" for reflected BB10.

cpar
: Copula parameter(s).

tau
: Kendall's tau.

Value

Kendall's tau or copula parameter.

References

Joe H (1997) *Multivariate Models and Dependence Concepts*. Chapman & Hall

Joe H (2014) *Dependence Modeling with Copulas*. Chapman & Hall/CRC.

Joe H (2014) *CopulaModel: Dependence Modeling with Copulas*. Software for book: *Dependence Modeling with Copulas*, Chapman & Hall/CRC, 2014.

Examples

```
# 1-param copulas
#BVN copula
cpar.bvn = tau2par("bvn", 0.5)
tau.bvn = par2tau("bvn", cpar.bvn)

#Frank copula
cpar.frk = tau2par("frk", 0.5)
tau.frk = par2tau("frk", cpar.frk)

#Gumbel copula
cpar.gum = tau2par("gum", 0.5)
tau.gum = par2tau("gum", cpar.gum)

#Joe copula
cpar.joe = tau2par("joe", 0.5)
tau.joe = par2tau("joe", cpar.joe)

# 2-param copulas
#BB1 copula
tau.bb1 = par2tau("bb1", c(0.5,1.5))

#BB7 copula
tau.bb7 = par2tau("bb7", c(1.5,1))
```

```
#BB8 copula
tau.bb8 = par2tau("bb8", c(3,0.8))

#BB10 copula
tau.bb10 = par2tau("bb10", c(3,0.8))
```

---

mle *Maximum likelihood estimation of factor copula models for mixed data*

Description

We use a two-stage etimation approach toward the estimation of factor copula models for mixed continuous and discrete data.

Usage

```
mle1factor(continuous, ordinal, count, copF1, gl, hessian, print.level)
mle2factor(continuous, ordinal, count, copF1, copF2, gl, hessian, print.level)
mle2factor.bvn(continuous, ordinal, count, copF1, copF2, gl, SpC, print.level)
```

Arguments

continuous	$n \times d_1$ matrix with the continuous reponse data, where n and d_1 is the number of observations and continous variables, respectively.
ordinal	$n \times d_2$ matrix with the ordinal reponse data, where n and d_2 is the number of observations and ordinal variables, respectively.
count	$n \times d_3$ matrix with the count reponse data, where n and d_3 is the number of observations and count variables, respectively.
copF1	$(d_1 + d_2 + d_3)$-vector with the names of bivariate copulas that link the each of the oberved variabels with the 1st factor. Choices are “bvn” for BVN, “bvtν” with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel, “joe” for Joe, “rjoe” for reflected Joe, “1rjoe” for 1-reflected Joe, “2rjoe” for 2-reflected Joe, “BB1” for BB1, “rBB1” for reflected BB1, “BB7” for BB7, “rBB7” for reflected BB7, “BB8” for BB8, “rBB8” for reflected BB8, “BB10” for BB10, “rBB10” for reflected BB10.
copF2	$(d_1 + d_2 + d_3)$-vector with the names of bivariate copulas that link the each of the oberved variabels with the 2nd factor. Choices are “bvn” for BVN, “bvtν” with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, “frk” for Frank, “gum” for Gumbel, “rgum” for reflected Gumbel, “1rgum” for 1-reflected Gumbel, “2rgum” for 2-reflected Gumbel, “joe” for Joe, “rjoe” for reflected Joe, “1rjoe” for 1-reflected Joe, “2rjoe” for 2-reflected Joe, “BB1” for BB1, “rBB1” for reflected BB1, “BB7” for BB7, “rBB7” for reflected BB7, “BB8” for BB8, “rBB8” for reflected BB8, “BB10” for BB10, “rBB10” for reflected BB10.

gl	Gauss legendre quardrature nodes and weights.
SpC	Special case for the 2-factor copula model with BVN copulas. Select a bivariate copula at the 2nd factor to be fixed to independence. e.g. "SpC = 1" to set the first copula at the 2nd factor to independence.
hessian	If TRUE, the hessian of the negative log-likelihood is calculated during the minimization process.
print.level	Determines the level of printing which is done during the minimization process; same as in nlm.

Details

Estimation is achieved by maximizing the joint log-likelihood over the copula parameters with the univariate parameters/distributions fixed as estimated at the first step of the proposed two-step estimation approach.

Value

A list containing the following components:

cutpoints	The estimated univariate cutpoints (fitting the univariate probit model).
negbinest	The estimated univariate parametes for the count responses (fitting the negative binomial distribution).
loglik	The maximized joint log-likelihood.
cpar	Estimated copula parameters in a list form.
taus	The estimated copula parameters in Kendall's tau scale.
SEs	The SEs of the Kendall's tau estimates.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References

Kadhem, S.H. and Nikoloulopoulos, A.K. (2019) Factor copula models for mixed data. *Arxiv e-prints*, <arXiv:1907.07395>. https://arxiv.org/abs/1907.07395.

Krupskii, P. and Joe, H. (2013) Factor copula models for multivariate data. *Journal of Multivariate Analysis*, **120**, 85–101. doi: 10.1016/j.jmva.2013.05.001.

Nikoloulopoulos, A.K. and Joe, H. (2015) Factor copula models with item response data. *Psychometrika*, **80**, 126–150. doi: 10.1007/s1133601393874.

Examples

```
#------------------------------------------------
# Setting quadreture points
nq <- 25
```

```
gl <- gauss.quad.prob(nq)
#-----------------------------------------------
#                    PE Data
#------------------             -----------------
data(PE)
continuous.PE1 = -PE[,1]
continuous.PE2 = PE[,2]
continuous.PE <- cbind(continuous.PE1, continuous.PE2)

categorical.PE <- PE[, 3:5]
#-----------------------------------------------
#                  Estimation
#------------------             -----------------
#------------------ One-factor  -----------------
# one-factor copula model
cop1f.PE <- c("joe", "joe", "rjoe", "joe", "gum")
est1factor.PE <- mle1factor(continuous.PE, categorical.PE,
                            count=NULL, copF1=cop1f.PE, gl, hessian = T)
est1factor.PE
#-----------------------------------------------
#-----------------------------------------------
#                    GSS Data
#------------------             -----------------
data(GSS)
attach(GSS)
continuous.GSS <- cbind(INCOME, AGE)
ordinal.GSS <- cbind(DEGREE, PINCOME, PDEGREE)
count.GSS <- cbind(CHILDREN, PCHILDREN)

#-----------------------------------------------
#                  Estimation
#------------------             -----------------
#------------------ One-factor  -----------------
# one-factor copula model
cop1f.GSS <- c("joe","2rjoe","bvt3","bvt3",
          "rgum","2rjoe","2rgum")
est1factor.GSS <- mle1factor(continuous.GSS, ordinal.GSS,
                        count.GSS, copF1 = cop1f.GSS, gl, hessian = T)

#------------------ Two-factor  -----------------
# two-factor copula model
cop1.2f <- c("rgum","rjoe","bvn","1rjoe",
           "1rjoe","rjoe","gum")
cop2.2f <- c("gum","2rjoe","rjoe","gum",
           "bvt5","bvn","2rgum")
est2factor.GSS <- mle2factor(continuous.GSS, ordinal.GSS,
          count.GSS, copF1 = cop1.2f, copF2 = cop2.2f, gl, hessian = T)
```

`PE` *Political-economic risk of 62 countries for the year 1987*

Description

Quinn (2004) used 5 mixed variables, namely the continuous variable black-market premium in each country (used as a proxy for illegal economic activity), the continuous variable productivity as measured by real gross domestic product per worker in 1985 international prices, the binary variable independence of the national judiciary (1 if the judiciary is judged to be independent and 0 otherwise), and the ordinal variables measuring the lack of expropriation risk and lack of corruption.

Usage

```
data(PE)
```

Format

A data frame with 62 observations (countries) on the following 5 variables:

`BM` Black-market premium.

`GDP` Gross domestic product.

`IJ` Independent judiciary.

`XPR` Lack of expropriation risk.

`CPR` Lack of corruption.

Source

Quinn, K. M. (2004). Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis*, **12**, 338–353.

`rfactor` *Simulation of factor copula models for mixed continuous and discrete data*

Description

Simulating dependent standard uniform and ordinal response data from factor copula models.

Usage

```
r1factor(n, d1, d2, categ, theta, copF1)
r2factor(n, d1, d2, categ, theta, delta, copF1, copF2)
```

Arguments

n	Sample size.
d1	Number of standard uniform variables.
d2	Number of ordinal variables.
categ	A vector of categories for the ordinal variables.
theta	Copula parameters for the 1st factor.
delta	Copula parameters for the 2nd factor.
copF1	$(d_1 + d_2)$-vector with the names of bivariate copulas that link the each of the oberved variabels with the 1st factor. Choices are "bvn" for BVN, "bvtν" with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, "frk" for Frank, "gum" for Gumbel, "rgum" for reflected Gumbel, "1rgum" for 1-reflected Gumbel, "2rgum" for 2-reflected Gumbel, "joe" for Joe, "rjoe" for reflected Joe, "1rjoe" for 1-reflected Joe, "2rjoe" for 2-reflected Joe, "BB1" for BB1, "rBB1" for reflected BB1, "BB7" for BB7, "rBB7" for reflected BB7, "BB8" for BB8, "rBB8" for reflected BB8, "BB10" for BB10, "rBB10" for reflected BB10.
copF2	$(d_1 + d_2)$-vector with the names of bivariate copulas that link the each of the oberved variabels with the 2nd factor. Choices are "bvn" for BVN, "bvtν" with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, "frk" for Frank, "gum" for Gumbel, "rgum" for reflected Gumbel, "1rgum" for 1-reflected Gumbel, "2rgum" for 2-reflected Gumbel, "joe" for Joe, "rjoe" for reflected Joe, "1rjoe" for 1-reflected Joe, "2rjoe" for 2-reflected Joe, "BB1" for BB1, "rBB1" for reflected BB1, "BB7" for BB7, "rBB7" for reflected BB7, "BB8" for BB8, "rBB8" for reflected BB8, "BB10" for BB10, "rBB10" for reflected BB10.

Value

Data matrix of dimension $n \times d$, where n is the sample size, and $d = d_1 + d_2$ is the total number of variables.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References

Kadhem, S.H. and Nikoloulopoulos, A.K. (2019) Factor copula models for mixed data. *Arxiv e-prints*, <arXiv:1907.07395>. https://arxiv.org/abs/1907.07395.

Examples

```
# ----------------------------------------------------
# ----------------------------------------------------
#             One-factor copula model
# ----------------------------------------------------
```

```
# ---------------------------------------------------
#Sample size ---------------------------------------
n = 100

#Continuous Variables  -----------------------------
d1 = 5

#Ordinal Variables  --------------------------------
d2 = 3

#Categories for ordinal  ---------------------------
categ = c(3,4,5)

#Copula parameters  --------------------------------
theta = rep(2, d1+d2)

#Copula names  -------------------------------------
copnamesF1 = rep("gum", d1+d2)

#----------------- Simulating data -----------------
datF1 = r1factor(n, d1=d1, d2=d2, categ, theta, copnamesF1)

#------------  Plotting continuous data -------------
pairs(qnorm(datF1[, 1:d1]))

# ---------------------------------------------------
# ---------------------------------------------------
#              Two-factor copula model
# ---------------------------------------------------
# ---------------------------------------------------
#Sample size ---------------------------------------
n = 100

#Continuous Variables  -----------------------------
d1 = 5

#Ordinal Variables  --------------------------------
d2 = 3

#Categories for ordinal  ---------------------------
categ = c(3,4,5)

#Copula parameters  --------------------------------
theta = rep(2.5, d1+d2)
delta = rep(1.5, d1+d2)

#Copula names  -------------------------------------
copnamesF1 = rep("gum", d1+d2)
copnamesF2 = rep("gum", d1+d2)

#----------------- Simulating data -----------------
datF2 = r2factor(n, d1=d1, d2=d2, categ, theta, delta,
                copnamesF1, copnamesF2)
```

```
#----------------  Plotting  data -----------------
pairs(qnorm(datF2[,1:d1]))
```

select *Model selection of the factor copula models for mixed data*

Description

An heuristic algorithm that automatically selects the bivariate parametric copula families that link the observed to the latent variables.

Usage

```
select1F(continuous, ordinal, count, copnamesF1, gl)
select2F(continuous, ordinal, count, copnamesF1, copnamesF2, gl)
```

Arguments

`continuous`	$n \times d_1$ matrix with the continuous reponse data, where n and d_1 is the number of observations and continous variables, respectively.
`ordinal`	$n \times d_2$ matrix with the ordinal reponse data, where n and d_2 is the number of observations and ordinal variables, respectively.
`count`	$n \times d_3$ matrix with the count reponse data, where n and d_3 is the number of observations and count variables, respectively.
`copnamesF1`	A vector with the names of possible candidates of bivariate copulas that link the each of the oberved variabels with the 1st factor. Choices are "bvn" for BVN, "bvtν" with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, "frk" for Frank, "gum" for Gumbel, "rgum" for reflected Gumbel, "1rgum" for 1-reflected Gumbel, "2rgum" for 2-reflected Gumbel, "joe" for Joe, "rjoe" for reflected Joe, "1rjoe" for 1-reflected Joe, "2rjoe" for 2-reflected Joe, "BB1" for BB1, "rBB1" for reflected BB1, "BB7" for BB7, "rBB7" for reflected BB7, "BB8" for BB8, "rBB8" for reflected BB8, "BB10" for BB10, "rBB10" for reflected BB10.
`copnamesF2`	A list with the names of possible candidates of bivariate copulas that link the each of the oberved variabels with the 1st and 2nd factors. Choices are "bvn" for BVN, "bvtν" with $\nu = \{1, \ldots, 9\}$ degrees of freedom for t-copula, "frk" for Frank, "gum" for Gumbel, "rgum" for reflected Gumbel, "1rgum" for 1-reflected Gumbel, "2rgum" for 2-reflected Gumbel, "joe" for Joe, "rjoe" for reflected Joe, "1rjoe" for 1-reflected Joe, "2rjoe" for 2-reflected Joe, "BB1" for BB1, "rBB1" for reflected BB1, "BB7" for BB7, "rBB7" for reflected BB7, "BB8" for BB8, "rBB8" for reflected BB8, "BB10" for BB10, "rBB10" for reflected BB10.
`gl`	Gauss legendre quardrature nodes and weights.

Details

The linking copulas at each factor are selected with a sequential algorithm under the initial assumption that linking copulas are Frank, and then sequentially copulas with non-tail quadrant independence are assigned to any of pairs where necessary to account for tail asymmetry (discrete data) or tail dependence (continuous data).

Value

A list containing the following components:

``1st factor'' The selected bivariate linking copulas for the 1st factor.

``2nd factor'' The selected bivariate linking copulas for the 2nd factor.

AIC Akaike information criterion.

taus The estimated copula parameters in Kendall's tau scale.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References

Kadhem, S.H. and Nikoloulopoulos, A.K. (2019) Factor copula models for mixed data. *Arxiv e-prints*, <arXiv:1907.07395>. https://arxiv.org/abs/1907.07395.

Examples

```
#-------------------------------------------------
#                   Estimation
#------------------             -----------------
# Setting quadreture points
nq<-25
gl<-gauss.quad.prob(nq)
#-------------------------------------------------
#                    PE Data
#------------------             -----------------
data(PE)
continuous.PE1 = -PE[,1]
continuous.PE <- cbind(continuous.PE1, PE[,2])
categorical.PE <- PE[, 3:5]

#------------------ One-factor  -----------------
# listing the possible copula candidates:
d <- ncol(PE)
copulasF1 <- rep(list(c("bvn", "bvt3", "bvt5", "frk", "gum",
"rgum", "rjoe","joe", "1rjoe","2rjoe", "1rgum","2rgum")), d)
out1F.PE <- select1F(continuous.PE, categorical.PE,
count=NULL, copulasF1, gl)
```

```
#-----------------------------------------------
#-----------------------------------------------
#                    GSS Data
#------------------             -----------------
data(GSS)
attach(GSS)
continuous.GSS <- cbind(INCOME, AGE)
ordinal.GSS <- cbind(DEGREE, PINCOME, PDEGREE)
count.GSS <- cbind(CHILDREN, PCHILDREN)

#------------------ One-factor  -----------------
# listing the possible copula candidates:
d <- ncol(GSS)
copulasF1 <- rep(list(c("bvn", "bvt3", "bvt5", "frk", "gum",
"rgum", "rjoe","joe", "1rjoe","2rjoe", "1rgum","2rgum")), d)
out1F.GSS <- select1F(continuous.GSS, ordinal.GSS, count.GSS, copulasF1, gl)

#------------------ two-factor  -----------------
# listing the possible copula candidates:
copulasF1 = copulasF2 = rep(list(c("bvn", "bvt3", "bvt5", "frk",
"gum", "rgum", "rjoe","joe", "1rjoe","2rjoe", "1rgum","2rgum")), d)
out2F.GSS <- select2F(continuous.GSS, ordinal.GSS,
count.GSS, copulasF1, copulasF2, gl)
```

`semicorr` *Diagnostics to detect tail dependence or tail asymmetry.*

Description

The sample versions of the correlation ρ_N, upper semi-correlation ρ_N^+ (correlation in the joint upper quadrant) and lower semi-correlation ρ_N^- (correlation in the joint lower quadrant). These are sample linear (when both variables are continuous), polychoric (when both variables are ordinal), and polyserial (when one variable is continuous and the other is ordinal) correlations.

Usage

```
semicorr(dat,type)
```

Arguments

`dat`	Data frame of mixed continuous and ordinal data.
`type`	A vector with 1's for the location of continuous data and 2's for the location of ordinal data.

Value

A matrix containing the following components for semicorr():

- `rho` ρ_N.
- `lrho` ρ_N^-.
- `urho` ρ_N^+.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References

Joe, H. (2014). *Dependence Modelling with Copulas*. Chapman and Hall/CRC.

Kadhem, S.H. and Nikoloulopoulos, A.K. (2019) Factor copula models for mixed data. *Arxiv e-prints*, <arXiv:1907.07395>. https://arxiv.org/abs/1907.07395.

Examples

```
#------------------------------------------------
#                      PE Data
#------------------             -----------------
data(PE)
#correlation
continuous.PE1 <- -PE[,1]
continuous.PE <- cbind(continuous.PE1, PE[,2])
categorical.PE <- data.frame(apply(PE[, 3:5], 2, factor))
nPE <- cbind(continuous.PE, categorical.PE)

#-------------------------------------------------
# Semi-correlations-------------------------------
#-------------------------------------------------
# Exclude the dichotomous variable
sem.PE = nPE[,-3]
semicorr.PE = semicorr(dat = sem.PE, type = c(1,1,2,2))
#------------------------------------------------
#------------------------------------------------
#                     GSS Data
#------------------             -----------------
data(GSS)
attach(GSS)
continuous.GSS <- cbind(INCOME,AGE)
ordinal.GSS <- cbind(DEGREE,PINCOME,PDEGREE)
count.GSS <- cbind(CHILDREN,PCHILDREN)

# Transforming the COUNT variables to ordinal
# count1 : CHILDREN
count1  = count.GSS[,1]
```

```
count1[count1 > 3] = 3

# count2: PCHILDREN
count2  = count.GSS[,2]
count2[count2 > 7] = 7

# Combining both transformed count variables
ncount.GSS = cbind(count1, count2)

# Combining ordinal and transformed count variables
categorical.GSS <- cbind(ordinal.GSS, ncount.GSS)
categorical.GSS <- data.frame(apply(categorical.GSS, 2, factor))

# combining continuous and categorical variables
nGSS = cbind(continuous.GSS, categorical.GSS)
#-------------------------------------------------
# Semi-correlations-------------------------------
#-------------------------------------------------
semicorr.GSS = semicorr(dat = nGSS, type = c(1, 1, rep(2,5)))
```

`transformation` *Continuous/count to ordinal responses*

Description

Transforming a continuous/count to ordinal variable with K categories.

Usage

```
continuous2ordinal(continuous, categ)
count2ordinal(count, categ)
```

Arguments

`continuous`	Matrix of continuous data.
`count`	Matrix of count data.
`categ`	The number of categories K.

Examples

```
#------------------------------------------------
#                     PE Data
#------------------              -----------------
data(PE)
```

```
continuous.PE <- PE[, 1:2]

#Transforming the continuous to ordinal data :
tcontinuous = continuous2ordinal(continuous.PE, categ=5)
table(tcontinuous)

#Transforming the count to ordinal data:
set.seed(12345)
count.data = rpois(1000, 3)
tcount = count2ordinal(count.data, 5)
table(tcount)
```

---

vuong *Vuong's test for the comparison of factor copula models*

Description

Vuong (1989)'s test for the comparison of non-nested factor copula models for mixed data. We compute the Vuong's test between the factor copula model with BVN copulas (that is the standard factor model) and a competing factor copula model to reveal if the latter provides better fit than the standard factor model.

Usage

```
vuong.1f(cpar.bvn, cpar, copF1, continuous, ordinal, count, gl, param)
vuong.2f(cpar.bvn, cpar, copF1, copF2, continuous, ordinal, count, gl, param)
```

Arguments

cpar.bvn	copula parameters of the factor copula model with BVN copulas.
cpar	copula parameters of the competing factor copula model.
copF1	copula names for the first factor of the competing factor copula model.
copF2	copula names for the second factor of the competing factor copula model.
continuous	matrix of continuous data.
ordinal	matrix of ordinal data.
count	matrix of count data.
gl	gauss-legendre quardature points.
param	parameterization of estimated copula parameters. If FALSE, then cpar are the actual copula parameters without any transformation/reparamterization.

Value

A vector containing the following components:

z	the test statistic.
p.value	the p-value.
CI.left	lower/left endpoint of 95% confidence interval.
CI.right	upper/right endpoint of 95% confidence interval.

Author(s)

Sayed H. Kadhem <s.kadhem@uea.ac.uk>
Aristidis K. Nikoloulopoulos <a.nikoloulopoulos@uea.ac.uk>

References

Kadhem, S.H. and Nikoloulopoulos, A.K. (2019) Factor copula models for mixed data. *Arxiv e-prints*, <arXiv:1907.07395>. https://arxiv.org/abs/1907.07395.

Vuong, Q.~H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**, 307–333.

Examples

```
#-------------------------------------------------
# Setting quadreture points
nq <- 25
gl <- gauss.quad.prob(nq)
#-------------------------------------------------
#                     PE Data
#------------------             -----------------
data(PE)
continuous.PE1 = -PE[,1]
continuous.PE2 = PE[,2]
continuous.PE <- cbind(continuous.PE1, continuous.PE2)
categorical.PE <- PE[, 3:5]
d <- ncol(PE)
#-------------------------------------------------
#                   Estimation
#------------------             -----------------
# factor copula model with BVN copulas
cop1f.PE.bvn <- rep("bvn", d)
PE.bvn1f <- mle1factor(continuous.PE, categorical.PE,
count=NULL, copF1=cop1f.PE.bvn, gl, hessian = T)

# Selected factor copula model
cop1f.PE <- c("joe", "joe", "rjoe", "joe", "gum")
PE.selected1f <- mle1factor(continuous.PE, categorical.PE,
count=NULL, copF1=cop1f.PE, gl, hessian = T)
#-------------------------------------------------
#                   Vuong's test
```

```
#------------------            -----------------
v1f.PE.selected <- vuong.1f(PE.bvn1f$cpar$f1,
PE.selected1f$cpar$f1,cop1f.PE, continuous.PE,
categorical.PE, count=NULL, gl, param=F)

#-----------------------------------------------
#-----------------------------------------------
#                    GSS Data
#------------------            -----------------
data(GSS)
attach(GSS)
continuous.GSS <- cbind(INCOME,AGE)
ordinal.GSS <- cbind(DEGREE,PINCOME,PDEGREE)
count.GSS <- cbind(CHILDREN,PCHILDREN)
d <- ncol(GSS)

#-----------------------------------------------
#                   Estimation
#------------------            -----------------
# factor copula model with BVN copulas
# one-factor copula model
cop1f.GSS.bvn <- rep("bvn", d)
GSS.bvn1f <- mle1factor(continuous.GSS, ordinal.GSS,
count.GSS, copF1 = cop1f.GSS.bvn, gl, hessian = T)

# two-factor copula model
cop1f.GSS.bvn = cop2f.GSS.bvn = rep("bvn", d)
GSS.bvn2f <- mle2factor.bvn(continuous.GSS, ordinal.GSS,
count.GSS, copF1 = cop1f.GSS.bvn, copF2 = cop2f.GSS.bvn, gl, SpC =7)

# Selected factor copula  model
# one-factor copula model
cop1f.GSS <- c("joe","2rjoe","bvt3","bvt3",
               "rgum","2rjoe","2rgum")
GSS.selected1f <- mle1factor(continuous.GSS, ordinal.GSS,
count.GSS, copF1 = cop1f.GSS, gl, hessian = T)

# two-factor copula model
cop2f1.GSS <- c("rgum","rjoe","bvn","1rjoe","1rjoe","rjoe","gum")
cop2f2.GSS <- c("gum","2rjoe","rjoe","gum","bvt5","bvn","2rgum")
GSS.selected2f <- mle2factor(continuous.GSS, ordinal.GSS,
count.GSS, copF1 = cop2f1.GSS, copF2 = cop2f2.GSS, gl, hessian = T)

#-----------------------------------------------
#                   Vuong's test
#------------------            -----------------
#1-factor
v1f.GSS.selected <- vuong.1f(GSS.bvn1f$cpar$f1,
GSS.selected1f$cpar$f1, cop1f.GSS, continuous.GSS,
ordinal.GSS, count=count.GSS, gl, param=F)

#2-factor
v2f.GSS.selected <- vuong.2f(GSS.bvn2f$cpar,
```

```
GSS.selected2f$cpar, cop2f1.GSS, cop2f2.GSS,
continuous.GSS, ordinal.GSS, count=count.GSS, gl, param=F)
```

Index