Package ‘causalreg’

March 12, 2025

Title Causal Generalized Linear Models

Version 0.1.0

Description An implementation of methods for causal discovery in a structural causal model where the conditional distribution of the target node is described by a generalized linear model conditional on its causal parents.

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.3

Imports mgcv

NeedsCompilation no

Author Veronica Vinciotti [aut, cre],
Ernst C. Wit [aut]

Maintainer Veronica Vinciotti `<veronica.vinciotti@unitn.it>`

Repository CRAN

Date/Publication 2025-03-12 17:30:20 UTC

Contents

`cgam` *Causal generalized additive model*

Description

This function does a search for a causal submodel within the generalized additive model provided.

Usage

```
cgam(
  formula,
  family,
  data,
  alpha = 0.05,
  pval.approx = FALSE,
  B = 100,
  seed = 1,
  search = c("all", "stepwise"),
  ...
)
```

Arguments

`formula`	A formula object.
`family`	A distributional family object. Currently supported options are: binomial and poisson.
`data`	A data frame containing the variables in the model.
`alpha`	Significance level for statistical test.
`pval.approx`	If TRUE, chi-squared approximated p-values are calculated. Default is FALSE, in which case p-values are calculated via bootstrap.
`B`	Number of bootstrap sample when pval.approx=FALSE.
`seed`	Seed for generating bootstrap samples.
`search`	If search="stepwise", a greedy forward stepwise search is conducted. Default is search="all", in which case all possible submodels are considered.
`...`	Further arguments to be passed to the gam function.

Value

A gam object of the selected causal submodel.

Examples

```
###############################
#causal Poisson gam##########
n<-1000
set.seed(123)
X1<-rnorm(n,0,1)
Y<-rpois(n,exp(sin(X1)))
X2<-log(Y+1)+rnorm(n,0,0.5)
data<-data.frame(cbind(X1, X2, Y))
cm_all<-cgam(Y ~ s(X1)+s(X2),"poisson",data,pval.approx=TRUE,search="all")
cm_all$model.opt
cm_step<-cgam(Y ~ s(X1)+s(X2),"poisson",data,pval.approx=TRUE,search="stepwise")
cm_step$model.opt
```

```
#bigger simulation with 7 covariates
set.seed(123)
n<-1000
X1<-rnorm(n=n,sd=sqrt(0.04))
X2<-X1+rnorm(n=n,sd=sqrt(0.04))
X3<-X1+X2+rnorm(n=n,sd=sqrt(0.04))
m<-sin(X2*5)+X3^3
Z<-m+rnorm(n=n,sd=sqrt(0.04))
X4<-X2+rnorm(n=n,sd=sqrt(0.04))
X5<-Z+rnorm(n=n,sd=sqrt(0.04))
X6<-Z+rnorm(n=n,sd=sqrt(0.04))
X7<-X6+rnorm(n=n,sd=sqrt(0.04))
Y<-qpois(pnorm(Z, mean = m, sd = sqrt(0.04)), lambda=exp(m))
dat<-data.frame(cbind(X1, X2, X3, X4, X5, X6, X7,Y))
fml<- Y~s(X1)+s(X2)+s(X3)+s(X4)+s(X5)+s(X6)+s(X7)
mod.all <-cgam(fml,"poisson",dat,pval.approx=TRUE,search="all")
mod.all$model.opt
mod.step <-cgam(fml,"poisson",dat,pval.approx=TRUE,search="stepwise")
mod.step$model.opt
####################################
#causal logistic gam################
n<-1000
set.seed(123)
X1<-rnorm(n,0,1)
Y<-rbinom(n,1,exp(X1)/(1+exp(X1)))
flip<-rbinom(n,1,0.1)
X2<-(1-flip)*Y+rnorm(n,0,0.3)
data<-data.frame(cbind(X1, X2, Y))
cm_all<-cgam(Y ~ s(X1)+s(X2),"binomial",data,pval.approx=FALSE,search="all")
cm_all$model.opt
cm_step<-cgam(Y ~ s(X1)+s(X2),"binomial",data,pval.approx=FALSE,search="stepwise")
cm_step$model.opt
```

`cglm` *Causal generalized linear model*

Description

This function does a search for a causal submodel within the generalized linear model provided.

Usage

```
cglm(
  formula,
  family,
  data,
  alpha = 0.05,
  pval.approx = FALSE,
  B = 100,
```

```
  seed = 1,
  search = c("all", "stepwise"),
  ...
)
```

Arguments

`formula`	A formula object.
`family`	A distributional family object. Currently supported options are: binomial and poisson.
`data`	A data frame containing the variables in the model.
`alpha`	Significance level for statistical test
`pval.approx`	If TRUE, chi-squared approximated p-values are calculated. Default is FALSE, in which case p-values are calculated via bootstrap.
`B`	Number of bootstrap sample when pval.approx=FALSE.
`seed`	Seed for generating bootstrap samples.
`search`	If search="stepwise", a greedy forward stepwise search is conducted. Default is search="all", in which case all possible submodels are considered.
`...`	Further arguments to be passed to the glm function.

Value

A glm object of the selected causal submodel.

Examples

```
###################################
#causal Poisson glm#################
n<-1000
set.seed(123)
X1<-rnorm(n,0,1)
Y<-rpois(n,exp(X1))
X2<-log(Y+1)+rnorm(n,0,0.3)
data<-data.frame(cbind(X1, X2, Y))
cm_all<-cglm(Y ~ X1+X2,"poisson",data,pval.approx=TRUE,search="all")
cm_all$model.opt
cm_step<-cglm(Y ~ X1+X2,"poisson",data,pval.approx=TRUE,search="stepwise")
cm_step$model.opt

##########################
#causal logistic glm#######
n<-2000
set.seed(123)
X1<-rnorm(n,0,1)
Y<-rbinom(n,1,exp(X1)/(1+exp(X1)))
flip<-rbinom(n,1,0.1)
X2<-(1-flip)*Y+rnorm(n,0,0.3)
data<-data.frame(cbind(X1, X2, Y))
cm_all<-cglm(Y ~ X1+X2,"binomial",data,pval.approx=FALSE,search="all")
```

```
cm_all$model.opt
cm_step<-cglm(Y ~ X1+X2,"binomial",data,pval.approx=FALSE,search="stepwise")
cm_step$model.opt
#bigger simulation with 5 covariates
set.seed(12)
n<-3000
X1<-rnorm(n,0,1)
X2<-rnorm(n,X1,0.5)
X3<-rnorm(n,0,1)
X4<-rnorm(n,X2,.5)
Y<-rbinom(n,1,exp(.8*X2-.9*X3)/(1+exp(.8*X2-.9*X3)))
flip<-rbinom(n,1,0.1)
X5<-(1-flip)*Y+flip*(1-Y)+rnorm(n,0,.3)
dat<-data.frame(cbind(X1, X2, X3, X4, X5,Y))
mod.all <-cglm(Y~X1+X2+X3+X4+X5,"binomial",dat,pval.approx=FALSE,search="all")
mod.all$model.opt
mod.step <-cglm(Y~X1+X2+X3+X4+X5,"binomial",dat,pval.approx=FALSE,search="stepwise")
mod.step$model.opt
```

Index