# Package 'MMGOR'

March 10, 2022

**Type** Package

**Title** Function for fitting the generalized odds rate model to clustered current status data using a novel minorize-maximize algorithm

**Version** 1.0

**Date** 2022-03-01

**Author** Tong Wang, Kejun He, Wei Ma, Dipankar Bandyopadhyay, Samiran Sinha

**Maintainer** Tong Wang <wangtong@nankai.edu.cn>

**Description** Parameter estimation and standard errors calculation for the generalized odds rate model with clustered current status data. The code can also handle the non-clustered data scenario. This package contains two functions. GOR_MM is used for analyzing data using the methodology proposed in Wang et al. SIM_DATA is used for simulating clustered/nonclustered current status data using the GOR model.

**License** Unlimited

**Imports** Rcpp (>= 1.0.4.6),gaussquad,extraDistr,fda,splines2,nleqslv,numDeriv

**Depends** gaussquad

**LinkingTo** Rcpp, RcppArmadillo

## R topics documented:



---

`GOR_MM` *Estimating parameters of the generalized odds rate model with clustered current status data.*

---

### Description

GOR_MM is used to analyze clustered current status data by fitting the generalized odds rate model. This is the main function of this package. It is assumed that a clustered dataset has $n$ units and there are $n_i$ subjects within the $i$th unit. Let $N = \sum_i n_i$ be the total number of subject. The response and explanatory variables are measured for every subjects. Response consists of $\Delta$, the censoring indicator, and $C$, the inspection time. It is assumed that there is also a set of unit level covariates.

### Usage

```
GOR_MM(Delta,X,Z,n,ni,r,C,knotsnum,order,cluster.ind=TRUE,
pen.ind=FALSE,lambda = 0,itermax=500,tol=1e-7,quadnum=30)
```

### Arguments

| | |
|---|---|
| `Delta` | A list of $n$ components, the $i$th component of the list is an $n_i \times 1$ vector containing indicators for censoring. The values 1 and 0 indicate left censored and right censored, respectively. |
| `X` | A list of $n$ components, the $i$th component is an $n_i \times p$ matrix of subject-level covariates. All columns of each matrix is assumed to be binary or numeric. For including a nominal categorical variable, the corresponding dummy variables must be included in this matrix. If there is no subject-level covariate (that means $p = 0$), then set $X$ =NULL. |
| `Z` | An $n \times q$ matrix of unit-level (cluster-level) covariates. All columns of this matrix is assumed to be binary or numeric. For including a nominal categorical variable, the corresponding dummy variables must be included in this matrix. If there is no unit-level covariate (that means $q = 0$), then set $Z$ =NULL. |
| `n` | An integer that denotes the number of units (clusters). |
| `ni` | An $n \times 1$ vector of integers. The sum of the elements of this vector equals to $N$. Each entry of this vector must be greater than one. If all $n$ entries are one, then it is no longer a clustered data, and cluster.ind is set to FALSE. |
| `r` | A scalar that specifies the generalized odds rate model. The value of $r$ must be non-negative. In this function $r = 0$ denotes the GOR model with $r \to 0$. |
| `C` | A list of $n$ components, the $i$th component is a $n_i \times 1$ vector containing the inspection time of all $n_i$ subjects within the $i$th unit. The inspection time must be positive for all subjects. |
| `knotsnum` | An integer indicating the number of equidistant interior knots for the integrated B-spline approximation of the nonparametric component of the GOR model. |
| `order` | An integer indicating the degree of integrated B-splines. |
| `cluster.ind` | Logical. If TRUE, the parameters will be estimated along with clustering effect. If it is FALSE, then all subjects are treated as independent and there is no clustering effect. The default value is TRUE. |
| `pen.ind` | Logical. If FALSE, there is no penalty term. The default value is FALSE. If TRUE, users should specify a positive value for the penalty term lambda. |
| `lambda` | A scalar containing the penalty parameter. The default value is 0. |
| `itermax` | The maximum number of iterations. The default value is 500. |
| `tol` | This denotes the sum of the absolute relative differences of the estimates at two successive iterations. It is used to check the convergence of the parameter estimates. The default value is 1e-7. |
| `quadnum` | The number of Gauss-Hermite quadrature nodes used in numerical integration. The default value is 30. |

### Value

Function GOR_MM returns a list containing the following components:

| | |
|---|---|
| result | The first column contains the estimate of the regression parameters. The second column contains the standard errors of these estimators. |
| AIC | The AIC value. |

## References

Wang, T., He, K., Ma, W., Bandyopadhyay, D., and Sinha, S. (2022). Minorize-maximize algorithm for the generalized odds rate model for clustered current status data. *Canadian Journal of Statistics*, in press.

## Examples

```
library(MMGOR)
set.seed(1)
H=function(t) log(1+t)+t^(3/2) # specify the H function

#This function generates a cohort with 300 subjects,
#where each subject has up to 8 observations.
#Both the subject-level covariates and unit-level covariates
# are generated from Uniform(-1,1) distribution.
#Users can use their own H function.

#Scenario 1
#3 subject-level covariates, 2 unit-level covariates (p=3,q=2)
data=SIM_DATA(r=0,beta=c(-1,1,0.5),gamma=c(-1,1),theta=1,n=300,H=H,up.quantile=0.85,
par_size=c(exp(1.7),1,8))
#estimate the parameters
result=GOR_MM(data$Delta,data$X,data$Z,data$n,data$ni,r=data$r,data$C,knotsnum = 2,order=2)

#Scenario 2
#only 2 unit-level covariates (p=0,q=2)
data=SIM_DATA(r=0,beta=c(0),gamma=c(-1,1),theta=1,n=300,H=H,up.quantile=0.85,
par_size=c(exp(1.7),1,8))
#estimate the parameters
result=GOR_MM(data$Delta,X=NULL,data$Z,data$n,data$ni,r=data$r,data$C,knotsnum = 2,order=2)

#Scenario 3
#only 3 subject-level covariates (p=3,q=0)
data=SIM_DATA(r=0,beta=c(-1,1,0.5),gamma=c(0),theta=1,n=300,H=H,up.quantile=0.85,
par_size=c(exp(1.7),1,8))
result=GOR_MM(data$Delta,data$X,Z=NULL,data$n,data$ni,r=data$r,data$C,knotsnum = 2,order=2)

#Scenario 4
#3 subject-level covariates, 2 unit-level covariates (p=3,q=2)
#no clustering effect
data=SIM_DATA(r=0,beta=c(-1,1,0.5),gamma=c(-1,1),theta=0,n=300,H=H,up.quantile=0.85,
par_size=c(exp(1.7),1,8))
result=GOR_MM(data$Delta,data$X,data$Z,data$n,data$ni,r=data$r,data$C,cluster.ind = FALSE,
knotsnum = 2,order=2)

# Scenario 5
# Data generation without the function SIM_DATA
n=100
ni= rpois(n, 4)
ni[ni<1]<-1
```

```
ni[ni>10]<-10
p=2;
q=1;
gamma=rep(-0.5, q)
beta=rep(0.5, p)
theta=0.8
xcov<-list()
zcov<-c()
mytime<-list()
mydelta<-list()
myinspection<-list()
zcov<-matrix(runif(n*q),nrow = n,ncol = q)

for( j in 1:n){
  xcov[[j]]<- matrix( runif((p*ni[j]), -1, 1), ncol=p )
  myrate= exp(as.numeric(zcov[j]%*%gamma)+
              as.numeric(xcov[[j]]%*%beta)-3+theta*rnorm(1))
  mytime[[j]]<-40+ rexp(ni[j],myrate)
  myinspection[[j]]<- runif(ni[j], 40, 90)
  mydelta[[j]]<- as.numeric(mytime[[j]]<myinspection[[j]] )
}
# End of data generation

### Analysis of the data using the GOR_MM function
result=GOR_MM(mydelta,xcov,zcov,n,ni,r=0,
              myinspection,knotsnum = 2,order=2)

### Analysis the same data with another r that is
### different from the data generating process.
result=GOR_MM(mydelta,xcov,zcov,n,ni,r=1.5,
myinspection,knotsnum = 2,order=2)
```

---

SIM_DATA *Generating datasets for the main function GOR_MM*

---

## Description

SIM_DATA is used to simulate clustered or non-clustered current status data. This function makes use of the generalized odds rate (GOR) model to generate the time-to-event.

## Usage

```
SIM_DATA(r, beta, gamma, theta, n, H, up.quantile,par_size)
```

## Arguments

| | |
|---|---|
| r | A scalar that specifies the generalized odds rate model. |
| beta | A $p \times 1$ vector representing the regression parameter for subject-level covariates. |
| gamma | A $q \times 1$ vector representing the regression parameter for unit-level covariates. |
| theta | A positive scalar representing the standard error of the clustering effect. When it is zero, the subjects are all independent (no clustering). |
| n | An integer that denotes the number of units (clusters). |

`H` A function representing the $H$ function included the generalized odds rate model. $H$ should be a non-negative and non-decreasing function with $H(0) = 0$.

`up.quantile` In this function, the inspection time is generated from a uniform distribution on $(0, b)$ with $b$ being the up.quantile% quantile of the simulated survival times. Users can change up.quantil within $(0, 1)$ to adjust the percentage of right censoring.

`par_size` In this function, the size of each unit is generated from a truncated Poisson distribution with the mean $\lambda$, and truncated below $\alpha$ and above $\beta$. Users should input the specific par_size=c($\lambda, \alpha, \beta$).

## Value

Function SIM_DATA returns a list containing the following components:

`X` A list of $n$ components, the $i$th component is an $n_i \times p$ matrix of subject-level covariates. All covariates are generated from Uniform$(-1, 1)$.

`Z` An $n \times q$ matrix of unit-level (cluster-level) covariates. All covariates are generated from Uniform$(-1, 1)$.

`n` An integer that denotes the number of units (clusters).

`ni` An $n \times 1$ vector of integers. Each elements of this vector represents the size of each unit.

`Delta` A list of $n$ components, the $i$th component of the list is an $n_i \times 1$ vector containing indicators for censoring. The values 1 and 0 indicate left censored and right censored, respectively.

`C` A list of $n$ components, the $i$th component is a $n_i \times 1$ vector containing the inspection time of all $n_i$ subjects within the $i$th unit. The inspection time must be positive for all subjects.

# Index