

Package ‘SPmlfcmcm’

December 11, 2015

Type Package

Version 1.4

Date 2015-12-10

Title Semiparametric Maximum Likelihood Method for Interactions
Gene-Environment in Case-Mother Control-Mother Designs

Author Moliere Nguile-Makao [aut, cre], Alexandre Bureau [aut]

Maintainer Moliere Nguile-Makao <moliere.nguile@gmail.com>

Depends nleqslv

Description Implements the method of general semiparametric maximum likelihood estimation for logistic models in case-mother control-mother designs.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2015-12-11 19:53:05

R topics documented:

| | |
|-----------------------------|---|
| SPmlfcmcm-package | 2 |
| Est.Inpar | 3 |
| FtSmlrmCMCM | 4 |
| SelteEch | 5 |
| Spmlfcmcm | 6 |

| | |
|--------------|-----------|
| Index | 10 |
|--------------|-----------|

| | |
|-------------------|------------------------------------------------------------------------------------------------|
| SPmlfcmcm-package | <i>SemiParametric Maximum Likelihood for interaction in case-mother control-mother designs</i> |
|-------------------|------------------------------------------------------------------------------------------------|

Description

Implementation of a method of general semiparametric maximum likelihood estimation for logistic models in case-mother control-mother designs. The method was proposed by *Chen et al., (2012)* for the complete data and *Nguile-Makao et al., (2015)* proposed an extension of the method allowing missing offspring genotype.

Details

The package **SPmlfcmcm** implements the semiparametric maximum likelihood estimation method published by *Chen et al., (2012)*. This method permits to analyze the interaction effects involving genetic variants and environmental exposures on the risk of adverse obstetric and early-life outcomes. *Nguile-Makao et al., (2015)* proposed an extension of this method allowing missing offspring genotype. The package performs the analysis the following way: it builds the nonlinear system from the database, resolves the nonlinear system using the `nleqslv` function of package **nleqslv**. It estimates the model parameters and the standard errors using the log profile likelihood function and the one-step method estimation. All this procedure may be done for complete data and also for missing offspring genotype. For more details see *Chen et al., (2012)*, and *Nguile-Makao et al., (2015)*. The modeling supposes that the distribution of maternal genotype and offspring genotype satisfy the following assumptions: *random mating*, *Hardy-Weinberg equilibrium* and *Mendelian inheritance*. The package also permits to treat the missing offspring genotype data.

Index of help topics:

| | |
|-------------------|-----------------------------------------------------------------------------------------|
| Est.Inpar | Computes the initial values |
| FtSmlrmCMM | Generates the logistic model data |
| SPmlfcmcm-package | SemiParametric Maximum Likelihood for interaction in case-mother control-mother designs |
| SeltcEch | Resampling |
| Spmlfcmcm | Semiparametric maximum likelihood for interaction in case-mother control-mother |

Author(s)

Moliere Nguile-Makao and Alexandre Bureau

Maintainer: Moliere Nguile-Makao <moliere.nguile@gmail.com>

References

Jinbo Chen, Dongyu Lin and Hagit Hochner (2012) Semiparametric Maximum Likelihood Methods for Analyzing Genetic and Environmental Effects with Case-Control Mother-Child Pair Data. *Biometrics* DOI: 10.1111/j.1541-0420.2011.01728.

Moliere Nguile-Makao, Alexandre Bureau (2015), Semi-Parametric Maximum likelihood Method for interaction in Case-Mother Control-Mother designs: Package SPmlfcmcm. *Journal of Statistical Software* DOI: 10.18637/jss.v068.i10.

See Also

[Est.Inpar](#), [FtSmIrmCMCM](#), [SeltcEch](#), [Spmlfcmcm](#)

| | |
|-----------|------------------------------------|
| Est.Inpar | <i>Computes the initial values</i> |
|-----------|------------------------------------|

Description

Computes initial values of the model parameters.

Usage

```
Est.Inpar(f1, N, gnma, gnch, tab1, typ, p = NULL)
```

Arguments

| | |
|------|--------------------------------------------------------------------------------------------------------------------|
| f1 | Model formula. |
| N | Numeric vector containing the number of eligible cases(N1) and controls (N0) in the population N = (N0, N1). |
| gnma | Name of mother genotype variable. |
| gnch | Name of offspring genotype variable. |
| tab1 | data.frame of the database. |
| typ | Argument that specifies whether the data are complete (1) or there are missing offspring genotypes (2). |
| p | Disease prevalence. |

Details

The function use logistic regression to evaluate the initial values of the equation parameters given by the formula and uses empirical estimation to compute the initial values of the nonlinear system. For details, see *Nguile-Makao et al., (2015)*.

Value

A list containing components

| | |
|-------|--------------------------------------------------------------------------------------------------|
| parms | Initial values of the model parameters given by the formula and the allelic frequency parameter. |
| ma.u | Initial values of nonlinear system. |

References

Moliere Nguile-Makao, Alexandre Bureau (2015). Semi-Parametric Maximum Likelihood Method for Interaction in Case-Mother Control-Mother Designs: Package SPmlfcmcm. *Journal of Statistical Software*, 68(10), 1-17. doi:10.18637/jss.v068.i10

FtSmlrmCMCM

Generates the logistic model data

Description

The function generates data from a logistic regression model. The data obtained contain: an outcome variable, the mother and child genotype coded as the number of minor allele and the environmental factors. For simulation of each environmental variable, the user can specify the coefficients of linear dependency between the mother genotype and the environmental factors.

Usage

```
FtSmlrmCMCM(f1, N, theta, beta, interc, vpo, vprob, vcorr)
```

Arguments

| | |
|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| f1 | Model formula. |
| N | Sample size. |
| theta | Minor allele frequency. |
| beta | Parameter vector of the effects. |
| interc | Intercept of the model. |
| vpo | Numeric vector containing the positions of the terms corresponding to the mother and child genotypes in the left-hand side of the formula. |
| vprob | Numeric vector containing the prevalence (success probability) of each environmental factor. |
| vcorr | Numeric vector containing the coefficients of linear dependency between the mother genotype and environmental factors. The value 0 corresponds to independence. |

Details

The function generates data, where the outcome variable is associated with the explanatory variables by a logistic regression model.

Ex: $\log(P/(1-P))=B_0+B_1*X_1+B_2*X_2+B_m*G_m+B_c*G_c+B_{2m}*X_2:G_m$.

Where $P=\Pr(Y=1|X)$, $X=(X_1,X_2)$ and Y is the outcome variable. The environmental factors are generated the following way: for each variable, a temporary variable is generated with a binomial law of success probability equal to $vprob[i]$ plus $vcorr[i]*G_m$, i is the factor position. The genotypes of the mother and her child are coded as the number of minor alleles, i.e. under an additive

model of the alleles on the log odds. The data generated suppose that the assumptions of *Hardy-Weinberg equilibrium*, *random mating type* and *Mendelian inheritance* are satisfied. The function uses the formula $f(x)=1/(1+\exp(-x))$ to generated the outcome variable. The data.frame returned by the function contains the variables whose names correspond to terms labels of the formula. The particularity of this function is to generate the genotype of a mother and her child taking into account the parental link.

Value

The function returns a data.frame containing an outcome variable, the environmental factors and two genotypes of the mother and her child.

Examples

```
# 1-Creation of database
set.seed(13200)
M=5000
f1=outc~X1+X2+gm+gc+X2:gm
vpo=c(3,4)
vprob=c(0.35,0.55)
vcorr=c(2,1)
theta=0.3;
beta=c(-0.916,0.857,0.405,-0.693,0.573)
interc=-2.23
Dataf<-FtSmlrmCMCM(f1,M,theta,beta,interc,vpo,vprob,vcorr)
Dataf[1:10,]
```

SeltcEch

Resampling

Description

The function draws randomly $n1+n0$ individuals among $N1 + N0$ individuals with $N0 > n0$ representing the numbers of non-cases and $N1 > n1$ representing the number of cases.

Usage

```
SeltcEch(outc, n1, n0, id, datf)
```

Arguments

| | |
|------|----------------------------------------------------|
| outc | Outcome variable (0,1). |
| n1 | Numeric value representing the number of cases. |
| n0 | Numeric value representing the number of controls. |
| id | Identifying number of the mother-child pair. |
| datf | data.frame of the database. |

Details

The function uses the [sample](#) function to resample the database.

Value

A data.frame with $n0 + n1$ rows.

See Also

[sample](#)

Examples

```
set.seed(13200)
M=5000;
f11=outc~Z1+Z2+Gm+Gc+Z2:Gm;
vpo=c(3,4)
vprob=c(0.35,0.55)
vcorr=c(2,1)
theta=0.3
beta=c(-0.916,0.857,0.405,-0.693,0.573)
interc=-2.23
Dataf<-FtSmlrmCMCM(f11,M,theta,beta,interc,vpo,vprob,vcorr)
# Number of subjects eligible to the study in the population
N0=dim(Dataf[Dataf["outc"]==0,])[1];
N1=dim(Dataf[Dataf["outc"]==1,])[1]
N=c(N0,N1)
# Sampling of the study database
n0=308
n1=83
DatfE1<-SeltcEch("outc",n1,n0,"obs",Dataf)
DatfE1[1:10,]
```

Spmlfcmcm

Semiparametric maximum likelihood for interaction in case-mother control-mother

Description

The function builds the nonlinear system from the data, solves the system and assesses the effect of each factor of the model, computes the variance - covariance matrix and deduces from it the standard deviations of each factor.

Usage

```
Spmlfcmcm(f1, N, gmname, gcname, DatfE, typ, start, p=NULL)
```

Arguments

| | |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| f1 | Model formula. |
| N | Numeric vector containing eligible number cases and controls in the study population $N=(N0, N1)$. |
| gmname | Name of mother genotype variable. |
| gcname | Name of offspring genotype variable. |
| DatfE | <code>data.frame</code> in long format containing the following variables: <i>outcome</i> variable, <i>mother genotype</i> , <i>offspring genotype</i> and <i>environmental factors</i> . |
| typ | Argument indicating whether the data are complete (1) or contain missing offspring genotypes (2). |
| start | Vector of the initial values of the model parameters. |
| p | Disease prevalence |

Details

The function `Spmlfcmcm` builds the nonlinear system from the data and solves the nonlinear system. Then, it uses the log profile likelihood function and the one-step method to estimate the parameters of each factor of the model formula and their standard errors. The programme computes the gradient of the profile likelihood using the analytical formula and the Hessian matrix numerically from the gradient. The genotype is coded as the number of minor alleles. The model supposes that the distribution of maternal genotype and offspring genotype satisfy the following assumptions: *random mating*, *Hardy-Weinberg equilibrium* and *Mendelian inheritance*. When the data contains missing offspring genotypes, the profile likelihood is summed over the possible genotypes of each child whose genotype is missing. The argument `typ` allows the user to specify whether the data is complete or not. Argument `start` permits to the user to give the initials values of model parameter. Ex: in the following equation $\log(P/(1-P))=B0+B1*X1+B2*X2+Bm*Gm+Bc*Gc+B2m*X2:Gm$, `start=(B0, B1, B2, Bm, Bc, B2m, fp)` where `fp` is the log of the odds of the minor allelic frequency. However, if the user provides no values, the function uses logistic regression to compute the initial $B=(B0, B1, B2, Bm, Bc, B2m)$ and takes 0.1 as the initial value of `fp`. If the argument `N` is unavailable, it is possible to specify the disease population prevalence in the argument `p` instead of `N`. In that case, `N1` is set equal to $5 n1$, in order to avoid observing $N1 < n1$ when prevalence is small. We then set $N0=[(1-p)/p]*N1$.

Value

A list containing components

| | |
|---------------|--------------------------------------------------------------------------------|
| Uim | Nonlinear system solution |
| MatR | Matrix containing the estimates and their standard errors |
| Matv | Variance - covariance matrix |
| Lhft | Log-likelihood function. It takes as argument a vector of the model parameters |
| Value_loglikh | Value of the Log-likelihood function computed at the parameters estimated |

References

Jinbo Chen, Dongyu Lin and Hagit Hochner (2012) Semiparametric Maximum Likelihood Methods for Analyzing Genetic and Environmental Effects with Case-Control Mother-Child Pair Data. *Biometrics* DOI: 10.1111/j.1541-0420.2011.01728.

Moliere Nguile-Makao, Alexandre Bureau (2015), Semi-Parametric Maximum likelihood Method for interaction in Case-Mother Control-Mother designs: Package SPmlfcmcm. *Journal of Statistical Software* DOI: 10.18637/jss.v068.i10.

Examples

```
# 1-Creation of database
## Not run:
  set.seed(13200)
  M=20000;
  fl=outc~X1+X2+gm+gnch+X1:gnch+X2:gm;
  theta=0.3
  beta=c(-0.916,0.857,0.588,0.405,-0.693,0.488)
  interc=-2.23
  vpo=c(3,4)
  vprob=c(0.35,0.55)
  vcorr=c(2,1)
  Dataf<-FtSmlrmCMCM(fl,M,theta,beta,interc,vpo,vprob,vcorr)
  rho<-table(Dataf$outc)[2]/20000 # Disease prevalence

# Number of subjects eligible to the study in the population
N=c(dim(Dataf[Dataf$outc==0,])[1],dim(Dataf[Dataf$outc==1,])[1])

# Sampling of the study database
n0=1232;n1=327;
DatfE1<-SeltcEch("outc",n1,n0,"obs",Dataf)

# 2 Creation of missing data on the offspring genotype
  DatfE=DatfE1
  gnch<-DatfE["gnch"]
  gnch<-as.vector(gnch[,1])
  gnch1<-sample(c(0,1),length(gnch),replace=TRUE,prob=c(0.91,0.09))
  gnch[gnch1==1]<-NA
  DatfE=DatfE1
  DatfE$gnch<-NULL;DatfE$gnch<-gnch
# 3 Creation of the two databases
  # DatfEcd :complete data
  # DatfEmd :data with missing genotypes for a subset of children.
  DatfEcd<-DatfE[is.na(DatfE["gnch"])!=TRUE,]
  DatfEmd<-DatfE
  rm(gnch);rm(gnch1)
# data obtained
DatfEcd[26:30,]
DatfEmd[26:30,]

##4 Estimation of parameters=====
```



```

## model equation
f1=outc~X1+X2+gm+gnch+X1:gnch+X2:gm;
## Estimation of the parameters (no missing data)
# N = (N0,N1) is available
Rsnm1<-Spmlfcmcm(f1,N,"gm","gnch",DatfEcd,1)
#solution of the nonlinear system
round(Rsnm1$Uim,digits=3)
#estimates
round(Rsnm1$MatR,digits=3)
#variance - covariance matrix
round(Rsnm1$Matv,digits=5)
# N = (N0,N1) is not available
Rsnm2<-Spmlfcmcm(f1=f1,gmname="gm",gcname="gnch",DatfE=DatfEcd,typ=1,p=rho)
#solution of the nonlinear system
round(Rsnm2$Uim,digits=3)
#estimates
round(Rsnm2$MatR,digits=3)
#variance - covariance matrix
round(Rsnm2$Matv,digits=5)
## Estimation of the parameters (with missing data)
# N = (N0,N1) is available
Rswm1<-Spmlfcmcm(f1,N,"gm","gnch",DatfEmd,typ=2)
#solution of the nonlinear system
round(Rswm1$Uim,digits=3)
#estimates
round(Rswm1$MatR,digits=3)
#variance - covariance matrix
round(Rswm1$Matv,digits=5)
# N = (N0,N1) is not available
Rswm2<-Spmlfcmcm(f1=f1,gmname="gm",gcname="gnch",DatfE=DatfEmd,typ=2,p=rho)
#solution of the nonlinear system
round(Rswm2$Uim,digits=3)
#estimates
round(Rswm2$MatR,digits=3)
#variance - covariance matrix
round(Rswm2$Matv,digits=5)

## End(Not run)

```

Index

- *Topic **Case-Control**
 - SPmlficmcm-package, 2
 - *Topic **Initial values**
 - Est.Inpar, 3
 - *Topic **Maximum Likelihood**
 - SPmlficmcm-package, 2
 - *Topic **Mother-Child Pair**
 - SPmlficmcm-package, 2
 - *Topic **Semiparametric**
 - SPmlficmcm-package, 2
 - *Topic **Simulation**
 - FtSmlrmCMCM, 4
 - *Topic **case-mother**
 - Spmlficmcm, 6
 - *Topic **maximum likelihood**
 - Spmlficmcm, 6
 - *Topic **semiparametric**
 - Spmlficmcm, 6
- Est.Inpar, 3, 3
- FtSmlrmCMCM, 3, 4
- nleqslv, 2
- sample, 6
- SeltcEch, 3, 5
- SPmlficmcm (SPmlficmcm-package), 2
- Spmlficmcm, 3, 6
- SPmlficmcm-package, 2