

Package ‘GxM’

September 5, 2014

Type Package

Title Maximum Likelihood Estimation for Gene-by-Measured Environment Interaction Models

Version 1.1

Date 2014-09-04

Author Hao Zheng and Paul Rathouz

Maintainer Hao Zheng <haozheng@stat.wisc.edu>

Description

Quantifying and testing gene-by-measured-environment interaction in behavior genetic designs.

License GPL-2

Imports methods, Rcpp, minqa, nlme, parallel

NeedsCompilation no

Repository CRAN

Date/Publication 2014-09-05 10:10:52

R topics documented:

checkGxM	2
GxM	3
summaryGxM	5

Index	7
--------------	----------

checkGxM

*Singularity Check for GxM Model Fittings***Description**

This function is used to check the fitted models. Model fittings based on numerical integration and derivative-free optimization may encounter circumstances in which likelihood reaches infinity because of singular covariance matrix or other issues. This function helps to detect such circumstances.

Usage

```
checkGxM(GxMmle, dataset, rGvalue=0.995,
         localpercentage=2, localrange=20, outlyingextent=2)
```

Arguments

GxMmle	a model fitting outcome from running GxM.
dataset	a data frame which is used to produce GxMmle.
rGvalue	rGvalue which is used to produce GxMmle.
localpercentage	a local percentage in terms of the value of parameter whose absolute value of gradient is the largest among all gradient components.
localrange	a positive number of parameter points with which the density values are computed and investigated.
outlyingextent	a threshold value measuring the outlying extent. A pair of twins/siblings would be identified if the variation of density values from two close set of parameters is larger than this threshold.

Value

locallikelihood	a vector of local log-likelihood values with only one varying parameter whose absolute value of gradient is the largest among all gradient components. The length of locallikelihood is equal to localrange.
localfMP	a data frame matrix of local density values corresponding to the largest gap of locallikelihood based on two items of localresponse and their difference. localresponse represents contribution from individual pairs of twins/siblings to the overall log-likelihood with the estimated parameters from GxMmle.
outlierID	an integer indicating the pair of twins/siblings contributing to the singularity. A returned value zero implies that no such pair of twins/siblings is found based on the current settings.

See Also

[GxM](#)

Examples

```
id <- 1:10;
rG <- c(rep(1,5),rep(0.5,5));
M1 <- rnorm(10);
M2 <- rnorm(10);
P1 <- rnorm(10);
P2 <- rnorm(10);
Data <- data.frame(id,rG,M1,M2,P1,P2);
output <- GxM(dataset = Data, modelName='Chol', closedform=TRUE);
checkGxM(output, dataset=Data);
```

GxM

Maximum Likelihood Estimation of Gene-by-Measured Environment Interaction Models

Description

This function is used to produce maximum likelihood estimation for model fittings. With identification of the model fitted to the given data set with certain constrains of parameters, both numerical method and closed form based method are available to obtain the maximum likelihood estimate of parameters. Hessian matrix of log-likelihood at the estimate is provided to obtain the standard errors of the estimate.

Usage

```
GxM(dataset, rGvalue=0.995, modelName,
     zeroset=character(), closedform=FALSE, K=8, coreNumber=1,
     manualinitial=NULL, priority=1, gradientlevel=2)
```

Arguments

dataset a data frame of input. This data frame should contain 6 columns, including the id of twin/sibling pairs, the rG value for the pairs, as well as the observed M1, M2, P1 and P2 values. The M1, M2, P1, P2 data would NOT be automatically standardized.

rGvalue a correlation value of genetic influences to a pair of twins/siblings. A value close to one is used to replace exact one to avoid calculations with singular matrices. This value cannot be the same for all pairs for the same reason.

modelName a character string input specifying which model should be fitted. Possible choices include Chol, CholGxM, CholNonLin, NLMainGxM, CorrGxM and CorrNonLin.

Chol $P = \mu_P + a_C * A_M + c_C * C_M + e_C * E_M + a_U * A_U + c_U * C_U + e_U * E_U.$

CholGxM $P = \mu_P + (a_C + \alpha_C * M) * A_M + (c_C + \kappa_C * M) * C_M + (e_C + \epsilon_C * M) * E_M + (a_U + \alpha_U * M) * A_U + (c_U + \kappa_U * M) * C_U + (e_U + \epsilon_U * M) * E_U.$

CholNonLin $P = \mu_P + a_C * A_M + c_C * C_M + e_C * E_M + a_U * A_U + c_U * C_U + e_U * E_U + \gamma_1 * A_M^2 + \gamma_2 * C_M^2 + \gamma_3 * E_M^2 + \delta_1 * A_M * A_U + \delta_2 * C_M * C_U + \delta_3 * E_M * E_U.$

	$\text{NLMainGxM } P = \mu_P + \beta_1 * M + \beta_2 * M^2 + (a_U + \alpha_U * M) * A_U + (c_U + \kappa_U * M) * C_U + (e_U + \epsilon_U * M) * E_U.$
	$\text{CorrGxM } P = \mu_P + (a_P + \alpha_P * M) * A_P + (c_P + \kappa_P * M) * C_P + (e_P + \epsilon_P * M) * E_P.$
	$r_A = \text{corr}(A_P, A_M); r_C = \text{corr}(C_P, C_M); r_E = \text{corr}(E_P, E_M)$
	$\text{CorrNonLin } P = \mu_P + a_P * A_P + c_P * C_P + e_P * E_P + \lambda_1 * A_M * A_P + \lambda_2 * C_M * C_P + \lambda_3 * E_M * E_P.$
	$r_A = \text{corr}(A_P, A_M); r_C = \text{corr}(C_P, C_M); r_E = \text{corr}(E_P, E_M)$
zeroset	a set of parameters that would be set zeros in the estimation. Character string for parameter names would be used for identification.
closedform	logical type indicating whether closed form method or numerical method should be applied to likelihood calculation.
K	an integer indicating the number of points in each dimension in numerical integration.
coreNumber	a positive integer representing whether and how parallel computation should be used. The fault value is set to be 1, and no parallel computation is necessary. A number greater than 1 indicates how many processors would be used in parallel computation.
manualinitial	a list of numerical values used for initializing the parameters excluding members in the zeroset.
priority	a positive number specifying the priority of manual initialization. It takes values in 1 and 2. Priority=2 implies that the manual initialization would override possible update built in the optimization.
gradientlevel	a positive value which serves as a threshold. A warning information would be produced if the maximum absolute value of gradient components exceeds this level.

Value

An S4 class of output with 10 slots would be returned.

loglikelihood	maximum log-likelihood value from fitting indicated model to the input data set.
BIC	Bayesian information criterion (BIC) value from the model fitting.
par	parameter estimates.
hess	Hessian matrix of log-likelihood at the maximum point with respect to parameters.
gradient	gradient components of log-likelihood at the maximum point with respect to parameters.
modelName	inherited from input.
zeroset	inherited from input.
closedform	inherited from input.
K	inherited from input.
coreNumber	inherited from input.

References

Rathouz, P. J., Van Hulle, C. A., Rodgers, J. L., Waldman, I. D., Lahey, B. B. (2008) *Specification, testing, and interpretation of gene-by-measured-environment interaction models in the presence of gene-environment correlation*. Behavior Genetics, 38(3), 301-315.

See Also

[checkGxM](#), [summaryGxM](#)

Examples

```
id <- 1:10;
rG <- c(rep(1,5),rep(0.5,5));
M1 <- rnorm(10);
M2 <- rnorm(10);
P1 <- rnorm(10);
P2 <- rnorm(10);
Data <- data.frame(id,rG,M1,M2,P1,P2);
output <- GxM(dataset = Data, modelname='Chol', closedform=TRUE);
output;
output@BIC;
output@hess;
```

summaryGxM

GxM Outcome Summaries

Description

This function is used to summarize the fitted models from applying function GxM.

Usage

```
summaryGxM(GxMmle, inverseHessian="overall", eps=1e-10, r=0.999)
```

Arguments

GxMmle	a model fitting outcome from running GxM.
inverseHessian	a character string specifying the way of inverting the Hessian matrix. It takes values in “overall” and “block”. The choice depends on whether the intercepts μ_M and μ_P should be treated disjointedly from other parameters. <code>inverseHessian=“overall”</code> inverts the whole Hessian matrix directly. <code>inverseHessian=“block”</code> chooses to invert two diagonal blocks of the Hessian matrix (a block corresponding to intercept parameters, and a block corresponding to non-intercept parameters), and combines the individual inverse.
eps	a boundary tolerance to see whether non-negative parameters reach zeros.
r	a boundary tolerance to see whether correlation coefficient parameters reach ones or negative ones.

Value

loglikelihood	maximum log-likelihood value from fitting indicated model to the input data set.
BIC	Bayesian information criterion (BIC) value from the model fitting.
par	parameter estimates.
stderr	standard errors of parameter estimates.

See Also[GxM](#)

Index

`checkGxM`, [2](#), [5](#)

`GxM`, [2](#), [3](#), [6](#)

`summaryGxM`, [5](#), [5](#)