Package 'PCH4GeneralPed'

June 30, 2015

Type Package

Title Principal Components of Heritability for High Dimension
Quantitative Traits and General Pedigrees

Version 1.0

Date 2012-07-29

Author Karim Oualkacha

Maintainer Karim Oualkacha <oualkacha.karim@uqam.ca>

Description package for the calculation of principal components of heritability for high dimension quantitative traits and general pedigrees.

License GPL (>= 2)

Depends kinship2

R topics documented:

PCH4GeneralPed-package

PCH4GeneralPed: R package for the calculation of principal components of heritability for high dimension quantitative traits and general pedigrees.

Description

PCH4GeneralPed: R package for the calculation of Principal Components of Heritability for High Dimension Quantitative Traits and General Pedigrees.

Details

This R package fits a linear mixed-effects model in the formulation described in Oualkacha et al. (2012) in the case of family-based designs and an arbitrary number of phenotypes. It controlls the family structure using a random effect and allows covariate effects as fixed effects. PCH4GeneralPed package estimates the genetic and the environmental variance components and the principal components of heritability based on an ANOVA approach. Fixed effects are estimated using generalized linear squares estimators.

The PCH4GeneralPed package uses the ridge penalized principal components of heritability proposed by Wang et al. (2007) to handle large number of traits.

Missing data are handled by eliminating the corresponding rows and columns from the analyzed data.

To see funcionality of the PCH4GeneralPed package, you can run the datasets examples: data(Ped.with.COVs) and data(Ped.without.COVs).

Note that except the main function "PCH4GeneralPed", all remaining functions are for personnel usage. Thus, users need to run only the main function "PCH4GeneralPed" to get principal components of heritability, see examples below.

Package:	PCH4GeneralPed
Type:	Package
Version:	1.0
Date:	2012-07-29
License:	(>= 2)

Author(s)

Karim Oualkacha et al.

Maintainer: Karim Oualkacha <oualkacha.karim@uqam.ca>

References

If you use PCH4GeneralPed package in your analysis, please cite the following work:

Oualkacha, K., Labbe, A., Ciampi, A., Roy, M.A. and Maziade, M., (2012). Principal components of heritability for high dimension quantitative traits and general pedigrees. Journal of Statistical Applications in Genetics and Molecular Biology, Volume 11. Issue 2, Article 4.

Wang Y, Fang Y, Jin M (2007). A ridge penalized principal-components approach based on heritability for high-dimensional data. Hum Hered, 64, 182-191.

Ott J, Rabinowitz D (1999). A principal-components approach based on heritability for combining phenotype information. Hum Hered, 49, 106-111.

Examples

```
data(Ped.with.COVs)
c1.COV = 6
c1.traits = 10
PCH4GeneralPed(Ped.with.COVs, c1.COV, c1.traits)

data(Ped.without.COVs)
c1.COV = 0
c1.traits = 6
PCH4GeneralPed(Ped.without.COVs, c1.COV, c1.traits)
```

ANOVA.data   *Transform pedigree data to a list format*

Description

This function transforms the pedigree file to a format which will be used by PCH4GeneralPed(.) function.

Usage

```
ANOVA.data(ntype, c1.traits, Ped)
```

Arguments

ntype	Vector of integers. Its length equals to the total number of families used in the pedigree file. The ntype's entries are the total number of subjects within each family. It can be obtained using Family.sizes(.) function.
c1.traits	Integer number indicates the column from which the traits begin in the Pedigree file.
Ped	Pedigree file with the first 5 columns should be in the following order: family IDs, subject IDs, father IDs, mother IDs and sex. If there are covariates in the model, they should be introduced after the five standard columns in the pedigree file followed by the quantitative traits.

Value

A list of matrices. Its length equals to the number of families used in the study. Each matrix of this list represents one family of the pedigree file (i. e. each list entry has subjects of a same family as rows and their corresponding traits as columns).

Author(s)

Karim Oualkacha

ANOVA.estimators	*Variance components estimators*

Description

This function estimates the genetic and the environmental variance components using an ANOVA approach.

Usage

```
ANOVA.estimators(ntype, Y, kin.miss)
```

Arguments

ntype
: Vector of integers. Its length equals to the total number of families used in the pedigree file. The ntype's entries are the total number of subjects within each family.

Y
: List of matrices obtained from the function ANOVA.data(). Its length equals to the number of families used in the study. Each matrix of this list represents one family of the pedigree file (i. e. each list entry has subjects of a same family as rows and their corresponding traits as columns).

kin.miss
: List of matrices. Its length equals to the total number of families. The list entries represent the family kinship matrices.

Value

nnd.Mnt.e
: Environmental variance-covariance matrix

nnd.Mnt.g
: Genetic variance-covariance matrix

SSW
: Subject-specific variance-covariance matrix

SSB
: family-specific variance-covariance matrix

Author(s)

Karim Oualkacha

References

Oualkacha, K., Labbe, A., Ciampi, A., Roy, M.A. and Maziade, M., (2012). Principal components of heritability for high dimension quantitative traits and general pedigrees. Journal of Statistical Applications in Genetics and Molecular Biology, Volume 11. Issue 2, Article 4.

ANOVA.estimators.COV   *Variance components estimates in the presence of covariates*

Description

This function estimates the genetic and the environmental variance components and the covariate effects using ANOVA and GLS approaches respectively.

Usage

```
ANOVA.estimators.COV(ntype, beta, X, Y, kin.miss)
```

Arguments

ntype
: Vector of integers. Its length equals to the total number of families used in the pedigree file. The ntype's entries are the total number of subjects within each family.

beta
: Ordinary least squares estimates of the covariate effects.

X
: Model design matrix.

Y
: List of matrices obtained from the function ANOVA.data(). Its length equals to the number of families used in the study. Each matrix of this list represents one family of the pedigree file (i. e. each list entry has subjects of a same family as rows and their corresponding traits as columns).

kin.miss
: kinship matrix as a list with length the total number of families. Eeach object of the list represents a family kinship matrix.

Value

nnd.Mnt.e
: Environmental variance-covariance matrix

nnd.Mnt.g
: Genetic variance-covariance matrix

SSW
: Subject-specific variance-covariance matrix

SSB
: family-specific variance-covariance matrix

Author(s)

Karim Oualkacha

References

Oualkacha, K., Labbe, A., Ciampi, A., Roy, M.A. and Maziade, M., (2012). Principal components of heritability for high dimension quantitative traits and general pedigrees. Journal of Statistical Applications in Genetics and Molecular Biology, Volume 11. Issue 2, Article 4.

---

bdiag   *Convert a list of matrices to one block diagonal matrix*

Description

This function converts a list of matrices to one block diagonal matrix.

Usage

```
bdiag(x)
```

Arguments

x   List of matrices to be converted to a one block diagonal matrix.

Value

Block diagonal matrix.

Author(s)

Karim Oualkacha

Coeff.beta.GLS   *Generalized Least Squares estimators of the covaraite effects*

Description

This function calculate GLS estimates of the covariate coefficients.

Usage

```
Coeff.beta.GLS(ntype, X, Y.Cov, Sigma.g, Sigma.e, kin.miss)
```

Arguments

ntype	Vector of integers. Its length equals to the total number of families used in the pedigree file. The ntype's entries are the total number of subjects within each family.
X	Model design matrix.
Y.Cov	Vector in which each subject is represented by the p columns (p is the number of traits considered in the study).
Sigma.g	An ANOVA estimate of the genetic variance component.
Sigma.e	An ANOVA estimate of the environmental variance component.
kin.miss	List of matrices. Its length equals to the total number of families. The list entries represent the family kinship matrices.

Value

Vector of the GLS estimates of the covariate effects, included the intercept.

Author(s)

Karim Oualkacha

CV.lambda.Wang   *Cross validation techniques for regularisation parameter*

Description

This function use cross validation techniques to calculate the regularisation parameter, Wang et al. (2007).

Usage

```
CV.lambda.Wang(lambda1, lambda2, ntype, N.part, u, Y, kin.miss)
```

Arguments

lambda1	Real number.
lambda2	Real number.
ntype	Vector of integers. Its length equals to the total number of families used in the pedigree file. The ntype's entries are the total number of subjects within each family.
N.part	Integer indicates the number of times that the data will be splited to training data and validation data.
u	Matrix of integers used to separate randomly the families to two groups of families: training and validation data.
Y	List of matrices obtained from the function ANOVA.data(). Its length equals to the number of families used in the study. Each matrix of this list represents one family of the pedigree file (i. e. each list entry has subjects of a same family as rows and their corresponding traits as columns).
kin.miss	List of matrices. Its length equals to the total number of families. The list entries represent the family kinship matrices.

Value

Cross validation heritability, see Wang et al. (2007).

Author(s)

Karim Oualkacha

References

Wang Y, Fang Y, Jin M (2007). A ridge penalized principal-components approach based on heritability for high-dimensional data. Hum Hered, 64, 182-191.

Data.Without.miss   *Get Pedigree data without missing values*

Description

This function keeps in the pedigree file only subjects without missing data.

Usage

```
Data.Without.miss(Ped)
```

Arguments

Ped   Pedigree file with the first 5 columns should be in the following order: family
               IDs, subject IDs, father IDs, mother IDs and sex. If there are covariates in the
               model, they should be introduced after the five standard columns in the pedigree
               file followed by the quantitative traits.

Value

A list of matrices. Its length equals to the number of families used in the study. Each matrix of this
list represents one family of the pedigree file (i. e. each list entry has subjects of a same family,
without missing data, as rows and their corresponding traits as columns).

Author(s)

Karim Oualkacha

DESIGN.Matrix.X   *Get the model design matrix*

Description

This function calculates the model design matrix from the pedigree file.

Usage

```
DESIGN.Matrix.X(p, ntype, c1.COV, c1.traits, Ped)
```

Arguments

p   The number of quantitative traits that are used in the pedigree file.

ntype   Vector of integers. Its length equals to the total number of families used in the
               pedigree file. The ntype's entries are the total number of subjects within each
               family.

c1.COV   Integer number indicates the column from which the covariates begin in the
               Pedigree file. If there is no covariate in the pedigree file, this parameter should
               be set to zero.

c1.traits	Integer number indicates the column from which the traits begin in the Pedigree file.
Ped	Pedigree file with the first 5 columns should be in the following order: family IDs, subject IDs, father IDs, mother IDs and sex. If there are covariates in the model, they should be introduced after the five standard columns in the pedigree file followed by the quantitative traits.

Value

The model design matrix.

Author(s)

Karim Oualkacha

Family.sizes	*Get family sizes from the pedigree file*

Description

This function calculates the family sizes of all the families of the pedigree file.

Usage

```
Family.sizes(Ped)
```

Arguments

Ped	Pedigree file, should have family IDs as first column

Value

Vector of length the number of families in the pedigree file. The vector entries are the number of subjects of the studied families.

Author(s)

Karim Oualkacha

kin.data	*Get kinship matrix*

Description

This function calculates the kinship matrix from a pedigree file.

Usage

```
kin.data(ntype, Ped)
```

Arguments

ntype   Vector of integers. Its length equals to the total number of families used in the
             pedigree file. The ntype's entries are the total number of subjects within each
             family.

Ped   Pedigree file with the first 5 columns should be in the following order: family
             IDs, subject IDs, father IDs, mother IDs and sex. If there are covariates in the
             model, they should be introduced after the five standard columns in the pedigree
             file followed by the quantitative traits.

Value

The kinship matrix as a list with length the total number of families. Eeach object of the list
represents a family kinship matrix.

Author(s)

Karim Oualkacha

See Also

Family.sizes

kin.data.miss	*Get kinship matrix for missing data*

Description

This function calculates the kinship matrix from a pedigree file in the presence of missing data.

Usage

```
kin.data.miss(ntype, Ped)
```

Arguments

ntype	Vector of integers. Its length equals to the total number of families used in the pedigree file. The ntype's entries are the total number of subjects within each family.
Ped	Pedigree file with the first 5 columns should be in the following order: family IDs, subject IDs, father IDs, mother IDs and sex. If there are covariates in the model, they should be introduced after the five standard columns in the pedigree file followed by the quantitative traits.

Value

List of matrices. Its length equals to the total number of families. The list entries represent the family kinship matrices.

Author(s)

Karim Oualkacha

See Also

Family.sizes, kin.data

PCH.Mment   *Principal components of heritability*

Description

This function calculates the ridge penalized principal components of heritability.

Usage

```
PCH.Mment(lambda, nnd.Mnt.e, nnd.Mnt)
```

Arguments

lambda	Regularization parameter (real number).
nnd.Mnt.e	The environmental variance component.
nnd.Mnt	The genetic variance component.

Value

Orth.Mment	Ridge penalized principal components of heritability
PCH.Mment1	First ridge penalized principal components of heritability
h.Mment	Vector of heritabilities of all the ridge penalized principal components
beta	Eigenvector of the generalized eigensystem problem
Diag.var.genetic	
	Eigenvalues of the generalized eigensystem problem, see the function solve-beta(.)

Author(s)

Karim Oualkacha

See Also

solvebeta

PCH4GeneralPed   *ANOVA Estimators of Principal Components of Heritability*

Description

Estimation of the principal components of heritability in family-based designs and an arbitrary number of phenotypes using an ANOVA approach.

Usage

```
PCH4GeneralPed(Ped, c1.COV, c1.traits)
```

Arguments

Ped	Pedigree file with the first 5 columns should be in the following order: family IDs, subject IDs, father IDs, mother IDs and sex. If there are covariates in the model, they should be introduced after the five standard columns in the pedigree file followed by the quantitative traits.
c1.COV	Integer indicates the column from which the covariates begin in the Pedigree file. If there is no covariate in the pedigree file, this parameter should be set to zero.
c1.traits	Integer indicates the column from which the traits begin in the Pedigree file.

Details

This function fits a linear mixed-effects model in the formulation described in Oualkacha et al. (2012) in the case of family-based designs and an arbitrary number of phenotypes. It controlls for the family structure using a random effect and allows covariate effects as fixed effects. PCH4GeneralPed function estimates the genetic and the environmental variance components and the principal components of heritability based on an ANOVA approach. Fixed effects are estimated using generalized linear squares estimators.

We use the ridge penalized principal components of heritability proposed by Wang et al. (2007) to handle large number of traits.

To see funcionality of the PCH4GeneralPed function, you can run the two datasets examples: data(Ped.with.COVs) and data(Ped.without.COVs).

Value

nbre.families	Integer number indicates the number of families in th pedigree file.
nbre.subjects.per.family	
	Vector of integers. Its length equals to the total number of families used in the pedigree file. The ntype's entries are the total number of subjects within each family.

```
Total.study.subjects
                Total subjects without missing data used in the study.
Intercept       Vector of the GLS estimate of the model intercept.
Coeff.Covariates
                Vector of the GLS estimates of the covariate effects.
Var.Polygenic   Genetic variance-covariance matrix.
Var.Env         Environmental variance-covariance matrix.
Corr.Polygenic  Genetic correlation matrix.
Corr.Env        Environmental correlation matrix.
All.PCH         Ridge penalized principal components of heritability.
Regularisation.parameter
                Regularisation parameter.
heritability.PCH
                Vector of heritabilities of all the ridge penalized principal components of heri-
                tability.
Trait.heritabilities
                Vector of single trait heritabilities.
Genetic.Variance.Proportion.PCH
                Vector with entries represent the proportion of the total genetic variation ex-
                plained by each principal component of heritability. The proportion of the total
                genetic variation explained by a gievn PCH, is defined as the ratio of the eigen-
                value of the genetic variance-covariance matrix that associated with this PCH
                and the sum of all the eigenvalues of the genetic variance-covariance matrix.
nbr.pch         Number of the significant principal components to be hold in the analysis.
```

Author(s)

Karim Oualkacha et al.

References

If you use PCH4GeneralPed package in your analysis, please cite the following work:

Oualkacha, K., Labbe, A., Ciampi, A., Roy, M.A. and Maziade, M., (2012). Principal components of heritability for high dimension quantitative traits and general pedigrees. Journal of Statistical Applications in Genetics and Molecular Biology, Volume 11. Issue 2, Article 4.

Wang Y, Fang Y, Jin M (2007). A ridge penalized principal-components approach based on heritability for high-dimensional data. Hum Hered, 64, 182-191.

Ott J, Rabinowitz D (1999). A principal-components approach based on heritability for combining phenotype information. Hum Hered, 49, 106-111.

Examples

```
data(Ped.with.COVs)
c1.COV = 6
c1.traits = 10
PCH4GeneralPed(Ped.with.COVs, c1.COV, c1.traits)

data(Ped.without.COVs)
c1.COV = 0
c1.traits = 6
PCH4GeneralPed(Ped.without.COVs, c1.COV, c1.traits)
```

Ped.with.COVs   *Pedigree file with covariates*

Description

Pedigree file with four covariates and five quantitative traits.

Usage

```
data(Ped.with.COVs)
```

Format

A data frame with 385 observations on the following 14 variables.

V1 Family IDs

V2 Subject IDs

V3 Father IDs

V4 Mother IDs

V5 Sex

V6 First covariate

V7 Second covariate

V8 Third covariate

V9 Forth covariate

V10 Trait 1

V11 Trait 2

V12 Trait 3

V13 Trait 4

V14 Trait 5

Details

This data contain 385 subjects (100 families) for which 5 correlated traits are simulated using a mixed model with a global mean alpha = (2,1,3,1,1) and 4 covariates with fixed effects beta = (1, 0.5, 2, 3).

Examples

```
data(Ped.with.COVs)

c1.COV = 6
c1.traits = 10
rand.seed.lambda = sample(1:1000,1)
PCH4GeneralPed(Ped.with.COVs, c1.COV, c1.traits, rand.seed.lambda)
```

Ped.without.COVs   *Pedigree file without covariates*

Description

Pedigree file with five quantitative traits and without covariates.

Usage

```
data(Ped.without.COVs)
```

Format

A data frame with 385 observations on the following 10 variables.

V1   Family IDs

V2   Subject IDs

V3   Father IDs

V4   Mother IDs

V5   Sex

V6   Trait 1

V7   Trait 2

V8   Trait 3

V9   Trait 4

V10   Trait 5

Details

This data contain 385 subjects (100 families) for which 5 correlated traits are simulated using a mixed model with a global mean alpha = (2,1,3,1,1) and without covariates.

Examples

```
data(Ped.without.COVs)

c1.COV = 0
c1.traits = 6
rand.seed.lambda = sample(1:1000,1)
PCH4GeneralPed(Ped.without.COVs, c1.COV, c1.traits, rand.seed.lambda)
```

---

solvebeta   *Generalized eigensystem problem*

Description

This function gives a solution of the generalized eigensystem problem, see Mardia et al. (1979).

Usage

```
solvebeta(lambda1, lambda2, SSw, SSb)
```

Arguments

lambda1	Real number.
lambda2	Real number.
SSw	Matrix semi-definite positive.
SSb	Matrix definite positive.

Value

beta	Eigenvector of the generalized eigensystem problem
d	Eigenvalues of the generalized eigensystem problem

Author(s)

Karim Oualkacha

References

Mardia KV, Kent JT, Bibby JM: Multivariate Analysis. London, Academic Press, 1979.

Y.cov   *Get all traits as one vector*

Description

This function transforms the traits from the pedigree file to a single vector

Usage

```
Y.cov(ntype, Y)
```

Arguments

ntype	Vector of integers. Its length equals to the total number of families used in the pedigree file. The ntype's entries are the total number of subjects within each family.
Y	List of matrices obtained from the function ANOVA.data(). Its length equals to the number of families used in the study. Each matrix of this list represents one family of the pedigree file (i. e. each list entry has subjects of a same family as rows and their corresponding traits as columns).

Value

vector in which each subject is represented by p columns, where p is the number of traits considered in the pedigree file.

Author(s)

Karim Oualkacha

Index