Package ‘coxKM’
October 12, 2015

Version 0.3
Date 2015-10-10
Title cox Kernel Machine SNP-set Association Test
Author Xinyi (Cindy) Lin, Qian Zhou
Maintainer Xinyi (Cindy) Lin <xinyilin@mail.harvard.edu>
Depends survival
Description SNP-set kernel association test for right-censored survival outcomes. coxKM is meant for common genetic variants only. coxKM tests for association between a SNP-set (made up of common variants) and a right-censored survival outcome.
License GPL (version 2 or later)

R topics documented:

<table>
<thead>
<tr>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>coxKM</td>
</tr>
<tr>
<td>examplecovariates</td>
</tr>
<tr>
<td>examplephenotype1</td>
</tr>
<tr>
<td>examplephenotype2</td>
</tr>
<tr>
<td>examplephenotype3</td>
</tr>
<tr>
<td>examplesnpset</td>
</tr>
</tbody>
</table>

Description

SNP-set kernel association test for right-censored survival outcomes.

Usage

coxKM(Z=NULL, U, Delta, X=NULL, gamma=NULL, kernel="linear", weights=NULL, npert=10^4, npert.check=TRUE, npert.upper=10^8, npert.threshold=50, impute.method = "fixed", is_check_genotype=TRUE, is_dosage=FALSE, missing_cutoff=0.15, SetID=NULL)
Arguments

\( X \)  

is a \( nxR \) matrix of relevant covariates with each row as a different individual and each column as a separate covariate measurement. If no additional covariates are present, \( X \) can be left unspecified or left as NULL. Note that each column of \( X \) has to be a numerical variable, non-numerical variables have to be recoded appropriately before analysis. \( X \) should not include an intercept.

\( Z \)  

is a \( nxS \) numeric genotype matrix with each row as a different individual and each column as a separate snp. Each genotype should be coded as 0, 1, 2, and 9 (or NA) for AA, Aa, aa, and missing, where A is a major allele and a is a minor allele. Missing genotypes will be imputed by the simple Hardy-Weinberg equilibrium (HWE) based imputation. If kernel matrix is supplied, \( Z \) is ignored and not used in testing.

\( U \)  

is a \( nx1 \) vector containing the observed times. Note: \( U=\min(C,T) \) where \( C = \) censoring time, \( T = \) survival time

\( Delta \)  

is a \( nx1 \) vector containing the status/event indicator.

\( gamma \)  

Unless \( X = \) NULL, \( gamma \) has to be supplied. \( gamma \) is the vector of coefficients from the null cox model corresponding to \( X \). \( gamma \) <- coxph(Surv(U,Delta)~X)$coef

\( kernel \)  

Type of kernel. \( kernel \) can be an \( nxn \) kernel matrix OR one of these six options: "linear.weighted", "linear", "IBS", "IBS.weighted", "quadratic" or "2wayIX". If an \( nxn \) kernel matrix is supplied, \( Z \) is ignored and is not used in testing.

\( weights \)  

is a vector of length \( S \) of prespecified weights for the weighted kernels. Weights in coxKM are defined the same way as in SKAT. The kernel matrix of the weighted linear kernel is \( K=ZWWZ' \).

\( npert \)  

is the number of perturbations used to calculate p-value (default =10000), \( npert \) should be at least 1000. Note that how small the p-value can be is limited by the number of perturbations. If \( npert.check = \) FALSE, \( npert \) perturbations to obtain an initial p-value and checks to see if the initial p-value <= \( npert.threshold/npert \). If the initial p-value <= \( npert.threshold/npert \), then \( npert.upper \) perturbations is used to obtain a more accurate p-value. Setting \( npert.check=TRUE \) allows a larger number of perturbations to be used to obtain more accurate p-values only when it is necessary. For very small p-values, it may be necessary to further increase \( npert.upper \).

\( npert.check \)  

TRUE/FALSE (default=TRUE). If \( npert.check=TRUE \), coxKM first uses \( npert \) perturbations to obtain an initial p-value and checks to see if the initial p-value <= \( npert.threshold/npert \). If the initial p-value <= \( npert.threshold/npert \), then \( npert.upper \) perturbations is used to obtain a more accurate p-value. Setting \( npert.check=TRUE \) allows a larger number of perturbations to be used to obtain more accurate p-values only when it is necessary. For very small p-values, it may be necessary to further increase \( npert.upper \).

\( npert.upper \)  

default=10^8. Used only if \( npert.check=TRUE \). See \( npert.check \).

\( npert.threshold \)  

default=50. Used only if \( npert.check=TRUE \). See \( npert.check \).

\( impute.method \)  

a method to impute missing genotypes (default= "fixed"). "random" imputes missing genotypes by generating binomial(2,p) random variables (p is the MAF), and "fixed" imputes missing genotypes by assigning the mean genotype value (2p). If you use "random", you will have different p-values for different runs because imputed values are randomly assigned. Can use set.seed() to replicate results.

\( is_check_genotype \)  

a logical value indicating whether to check the validity of the genotype matrix Z (default= TRUE). If you use non-SNP type data and want to run coxKM, please
set it to FALSE. If you use SNP data or imputed data, please set it to TRUE. If
is_check_genotype=FALSE, missing values in Z have to be coded only as NA
since 9 will not be treated as a missing value.

is_dosage  a logical value indicating whether the matrix Z is a dosage matrix (default=
FALSE). If is_dosage=TRUE, “is_check_genotype” and “impute.method” will
be ignored and coxKM will check the genotype matrix and set impute.method=“fixed”.  
Note that coxKM will also treat 9 as missing in Z.

missing_cutoff a cutoff of the missing rates of SNPs (default=0.15). Any SNPs with missing
rates higher than cutoff will be excluded from the analysis.

SetID  SetID.

Details  
If kernel is not a matrix and Z is supplied, and either is_check_genotype=TRUE OR is_dosage=TRUE,
coxKM will check the Z matrix for missing values (missing values must be coded either as NA or 9)
and apply imputation. If you are using coxKM for non-SNP/dosage data, set is_check_genotype=FALSE
and is_dosage=FALSE, in which case missing values must be coded as NA (9 is not considered a
missing value).

Value  
p.value  the p-value of coxKM based on resampling. Note that if the p-value takes on
the smallest possible value based on the number of perturbations, it may be
necessary to increase npert and npert.upper. See npert.check.

Q  the unscaled score test statistic of coxKM.

n.marker.test no. of SNPs used for testing, <=S.
n.indiv  n = no. of samples

df  the estimated degrees of freedom of the test statistic (for reference only, not used
in association testing).

Author(s)  
Xinyi (Cindy) Lin, Qian Zhou

References

SNP-set Analysis for Genome-wide Association Studies. Genetic Epidemiology 35:620-31. doi:
10.1002/gepi.20610

Cai T, Tonini G and Lin X. 2011. Kernel machine approach to testing the significance of multiple

Examples

data(examplesnpset, examplecovariates, examplephenotype1, examplephenotype2, examplephenotype3)
Z <- as.matrix(examplesnpset)
X <- as.matrix(examplecovariates)
phenotype1 <- examplephenotype1
examplephenotype1

```r
phenotype2 <- examplephenotype2
phenotype3 <- examplephenotype3
set.seed(1)

#-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
# coxKM without covariates
#-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
coxKM(Z=Z, U=phenotype1$time, Delta=phenotype1$event, kernel="IBS")
coxKM(Z=Z, U=phenotype1$time, Delta=phenotype1$event, kernel="linear")
coxKM(Z=Z, U=phenotype3$time, Delta=phenotype3$event, kernel="IBS")
coxKM(Z=Z, U=phenotype3$time, Delta=phenotype3$event, kernel="linear")

#-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
# coxKM with covariates
#-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Gamma <- coxph(Surv(phenotype2$time, phenotype2$event)~X)$coef
Gamma
coxKM(Z=Z, U=phenotype2$time, Delta=phenotype2$event, X=X, gamma=Gamma, kernel="IBS")
coxKM(Z=Z, U=phenotype2$time, Delta=phenotype2$event, X=X, gamma=Gamma, kernel="linear")
```

descriptions:

`examplecovariates` *Example covariates dataset for coxKM.*

**Description**

Example covariates dataset for coxKM.

**Format**

eexamplecovariates contains:

- a numeric matrix of 2000 individuals and 2 covariates. Each row represents a different individual. coxKM.examplecovariates is identical to X in SKAT.example.

**Author(s)**

Xinyi (Cindy) Lin

descriptions:

`examplephenotype1` *Example phenotype for coxKM.*

**Description**

Example phenotype for coxKM.

**Format**

eexamplephenotype1 contains:

- a numeric matrix with two columns, the first column is the event time, the second column is the status indicator. Each row represents a different individual.
examplephenotype2

Description
Example phenotype for coxKM.

Format
examplephenotype2 contains:
- a numeric matrix with two columns, the first column is the event time, the second column is the status indicator. Each row represents a different individual.

Author(s)
Xinyi (Cindy) Lin

examplephenotype3

Description
Example phenotype for coxKM.

Format
examplephenotype3 contains:
- a numeric matrix with two columns, the first column is the event time, the second column is the status indicator. Each row represents a different individual.

Author(s)
Xinyi (Cindy) Lin
Example SNP-set for coxKM.

**Description**
Example SNP-set for coxKM.

**Format**
examplesnpset contains:
- a numeric genotype matrix of 2000 individuals and 11 SNPs. Each row represents a different individual, and each column represents a different SNP marker. coxKM.examplesnpset is subset of Z in SKAT.example.

**Author(s)**
Xinyi (Cindy) Lin
Index

coxKM, 1
examplecovariates, 4
examplephenotype1, 4
examplephenotype2, 5
examplephenotype3, 5
examplesnpset, 6