

The SNPmaxsel Package

February 16, 2008

Version 1.0-1

Date 2007-06-11

Title Maximally selected statistics for SNP data

Author Anne-Laure Boulesteix <boulesteix@slcmsr.org>.

Maintainer Anne-Laure Boulesteix <boulesteix@slcmsr.org>

Depends R (>= 2.2.1), mvtnorm, combinat

Suggests mvtnorm, combinat

Description This package implements asymptotic methods related to maximally selected statistics, with applications to SNP data.

License GPL version 2 or newer

URL <http://cran.r-project.org/src/contrib/Descriptions/SNPmaxsel.html>

R topics documented:

Fasympt	2
SNPmaxsel-internal	3
maxsel.asymp.test	3
maxsel	5
transf.inter	6
Index	8

Fasymp

*Asymptotical distribution of the maximally selected chi-square statistic***Description**

This function computes the asymptotical distribution of the maximally selected chi-square statistic, where maximal selection is performed over the test statistics defined by `type` or user-defined through `groups`.

Usage

```
Fasymp(t, a.vec, type=NULL, groups=NULL)
```

Arguments

<code>t</code>	the value at which the distribution function has to be computed.
<code>a.vec</code>	A vector of length <code>K</code> giving the proportion of observations in each category. Its sum must be 1. If <code>type="inter.ord"</code> or <code>"inter.cat"</code> , <code>a.vec</code> must have length 9.
<code>type</code>	must be one of <code>"ordinal"</code> , <code>"all.pairs"</code> , <code>"all.partitions"</code> , <code>"inter.ord"</code> , <code>"inter.cat"</code> , <code>"inter.ord.main"</code> . If <code>type=NULL</code> , the chi-square statistics are computed for the groups defined by <code>groups</code>
<code>groups</code>	If <code>type=NULL</code> , the chi-square statistics are computed for the groups defined by <code>groups</code> . <code>groups</code> must be a list with <code>m</code> elements, where <code>m</code> is the number of considered chi-square statistics. Each element is a list with two elements <code>group1</code> and <code>group2</code> . <code>groups\$group1</code> and <code>groups\$group2</code> are numeric vectors giving the indices of the categories included in both groups. See example below.

Details

This function uses the function `pmvnorm` from the package `mvtnorm`. If the considered test statistics are defined by `groups`, `groups` should not have more than 100 elements, since the `pmvnorm` function becomes unstable (or computationally prohibitive) for such a high dimension.

Value

the value of the distribution function at `t`.

Author(s)

Anne-Laure Boulesteix (<http://www.slcmr.net/boulesteix>)

References

Boulesteix AL, Strobl C, Weidinger S, Wichmann HE, Wagenpfeil S, 2007. Multiple testing for SNP-SNP interactions. Submitted.

See Also

`maxsel.asymp.test`, `maxsel`.

Examples

```
# load SNPmaxsel library
# library(SNPmaxsel)

Fasymp(t=2, a.vec=c(0.2, 0.1, 0.05, 0.05, 0.6), type="ord")
Fasymp(t=2, a.vec=c(0.2, 0.1, 0.05, 0.05, 0.6), type="all.pairs")
Fasymp(t=2, a.vec=c(0.2, 0.1, 0.05, 0.05, 0.6), type="all.partitions")

Fasymp(t=2, a.vec=c(0.2, 0.1, 0.1, 0.1, 0.05, 0.05, 0.05, 0.05, 0.3), type="inter.ord")
Fasymp(t=2, a.vec=c(0.2, 0.1, 0.1, 0.1, 0.05, 0.05, 0.05, 0.05, 0.3), type="inter.cat")

# Creating a user-defined this-is-escaped-code{ list
# (corresponding to this-is-escaped-codenormal-bracket47bracket-normal)="ord" with K=5)
my.groups<-list(list(group1=1, group2=2:5), list(group1=1:2, group2=3:5),
list(group1=1:3, group2=4:5), list(group1=1:4, group2=5))

Fasymp(t=2, a.vec=c(0.2, 0.1, 0.05, 0.05, 0.6), type=NULL, groups=my.groups)
```

SNPmaxsel-internal *Internal SNPmaxsel Functions*

Description

Internal SNPmaxsel functions.

Note

These are not to be called by the user (or in some cases are just waiting for proper documentation to be written).

`maxsel.asymp.test` *Test of independence based on maximally selected statistics*

Description

The function `maxsel.test` computes the probability that the maximally selected chi-square statistic is \leq than the value observed from the data, under the null-hypothesis of no association between X and Y , given the proportions of observations with $X=1, \dots, X=K$. The candidate binary splits over which the chi-square statistic is maximized is specified by `type`. If x denotes the output of the function `maxsel.asymp.test`, $1-x$ may be seen as the p-value of an independence test.

Usage

```
maxsel.asymp.test(x1, x2=NULL, y, type)
```

Arguments

x1	a numeric vector of length n giving the values of the variable x1 for the considered n observations. The classes must be coded as 1,...,K.
x2	a numeric vector of length n giving the values of the variable x2 for the considered n observations. x2 should be NULL (default) for all types other than "inter.ord", "inter.cat" and "inter.ord.main". Since interactions are implemented for K=3 only (SNPs), x2 must be coded as 1,2,3.
y	a numeric vector of length n giving the class (response variable Y) of the considered n observations. The classes must be coded as 0 and 1.
type	must be one of "ordinal", "all.pairs", "all.partitions", "inter.ord", "inter.cat", "inter.ord.main".

Details

See Boulesteix et al (2007).

Value

maxselstat	the observed maximally selected statistic.
value	the value of the distribution function of the maximally selected statistic at maxselstat.

Author(s)

Anne-Laure Boulesteix (<http://www.slcmr.net/boulesteix>)

References

Boulesteix AL, Strobl C, Weidinger S, Wichmann HE, Wagenpfeil S, 2007. Multiple testing for SNP-SNP interactions. Submitted.

See Also

[maxsel](#), [Fasymp](#).

Examples

```
# load SNPmaxsel library
# library(SNPmaxsel)

x1<-sample(5,1000,replace=TRUE)
y<-sample(c(0,1),1000,replace=TRUE)

maxsel.asymp.test(x1=x1,y=y,type="ord")
maxsel.asymp.test(x1=x1,y=y,type="all.pairs")
maxsel.asymp.test(x1=x1,y=y,type="all.partitions")
```

```
x1<-sample(3,1000,replace=TRUE)
x2<-sample(3,1000,replace=TRUE)

maxsel.asymp.test(x1=x1,x2=x2,y=y,type="inter.ord")
maxsel.asymp.test(x1=x1,x2=x2,y=y,type="inter.cat")
maxsel.asymp.test(x1=x1,x2=x2,y=y,type="inter.ord.main")
```

`maxsel`*Computes maximally selected chi-square statistics*

Description

The function `maxsel` computes the maximal chi-square statistic over some candidate binary splits specified by `type`.

Usage

```
maxsel(x1,x2=NULL,y,type="inter.ord")
```

Arguments

<code>x1</code>	a numeric vector of length <code>n</code> giving the values of the variable <code>x1</code> for the considered <code>n</code> observations. The classes must be coded as 1,...,K.
<code>x2</code>	a numeric vector of length <code>n</code> giving the values of the variable <code>x2</code> for the considered <code>n</code> observations. <code>x2</code> should be <code>NULL</code> (default) for all types other than <code>"inter.ord"</code> , <code>"inter.cat"</code> and <code>"inter.ord.main"</code> . Since interactions are implemented for <code>K=3</code> only (SNPs), <code>x2</code> must be coded as 1,2,3.
<code>y</code>	a numeric vector of length <code>n</code> giving the class (response variable <code>Y</code>) of the considered <code>n</code> observations. The classes must be coded as 0 and 1.
<code>type</code>	must be one of <code>"ordinal"</code> , <code>"all.pairs"</code> , <code>"all.partitions"</code> , <code>"inter.ord"</code> , <code>"inter.cat"</code> , <code>"inter.ord.main"</code> .

Details

Value

the value of the maximally selected chi-square statistic.

Author(s)

Anne-Laure Boulesteix (<http://www.slcmsr.net/boulesteix>)

References

Boulesteix AL, Strobl C, Weidinger S, Wichmann HE, Wagenpfeil S, 2007. Multiple testing for SNP-SNP interactions. Submitted.

See Also

[maxsel.asymp.test](#), [Fasymp](#).

Examples

```
# load SNPmaxsel library
# library(SNPmaxsel)

x1<-sample(5,1000,replace=TRUE)
y<-sample(c(0,1),1000,replace=TRUE)

maxsel(x1=x1,y=y,type="ord")
maxsel(x1=x1,y=y,type="all.pairs")
maxsel(x1=x1,y=y,type="all.partitions")

x1<-sample(3,1000,replace=TRUE)
x2<-sample(3,1000,replace=TRUE)

maxsel(x1=x1,x2=x2,y=y,type="inter.ord")
maxsel(x1=x1,x2=x2,y=y,type="inter.cat")
maxsel(x1=x1,x2=x2,y=y,type="inter.ord.main")
```

transf.inter

Transforms a pair of SNPs into a single variable with nine categories

Description

The function `transf.inter` creates a new variable with 9 categories out of a pair of variables with 3 categories (e.g. SNPs). The new variable is generated as described in Boulesteix et al (2007): =1 if $x_1=1$ and $x_2=1$, =2 if $x_1=2$ and $x_2=1$, =3 if $x_1=3$ and $x_2=1$, =4 if $x_1=1$ and $x_2=2$, =5 if $x_1=2$ and $x_2=2$, =6 if $x_1=3$ and $x_2=2$, =7 if $x_1=1$ and $x_2=3$, =8 if $x_1=2$ and $x_2=3$, =9 if $x_1=3$ and $x_2=3$.

Usage

```
transf.inter(x1,x2)
```

Arguments

`x1` a numeric vector of length `n` giving the first SNP, coded as 1,2,3.
`x2` a numeric vector of length `n` giving the second SNP, coded as 1,2,3.

Details

Value

a numeric vector of length n containing the new variable with 9 categories.

Author(s)

Anne-Laure Boulesteix (<http://www.slcmr.net/boulesteix>)

References

Boulesteix AL, Strobl C, Weidinger S, Wichmann HE, Wagenpfeil S, 2007. Multiple testing for SNP-SNP interactions. Submitted.

See Also

[maxsel](#), [maxsel.asymp.test](#).

Examples

```
# load SNPmaxsel library
# library(SNPmaxsel)

x1<-sample(3,100,replace=TRUE)
x2<-sample(3,100,replace=TRUE)

transf.inter(x1,x2)
```

Index

*Topic **h**test

- Fasymp, 1
- maxsel, 5
- maxsel.asymp.test, 3
- transf.inter, 6

*Topic **i**nternal

- SNPmaxsel-internal, 3

chisq.test.a
(*SNPmaxsel-internal*), 3

Fasymp, 1, 4, 5

groups (*SNPmaxsel-internal*), 3

makeA (*SNPmaxsel-internal*), 3

makeB (*SNPmaxsel-internal*), 3

maxsel, 2, 4, 5, 7

maxsel.asymp.test, 2, 3, 5, 7

SNPmaxsel-internal, 3

transf.inter, 6