Package 'jointsum'

October 29, 2018

Title jointsum

Version 0.19

Author Yangqing Deng [aut, cre],
Wei Pan [aut]

Maintainer Yangqing Deng <yangq001@umn.edu>

Date 2018-01-11

Description Joint analysis using summary statistics. Build joint models. Test for pleiotropy or allelic heterogeneity.

Depends R (>= 3.2.4)

Imports MASS

License MIT

Encoding UTF-8

LazyData TRUE

RoxygenNote 6.0.1.9000

R topics documented:

---

conAH *conAH*

Description

Test Allelic Heterogeneity with summary statistics and intersection-union tests.

Usage

```
conAH(Z,ld,n)
```

Arguments

Z	a q*1 matrix containing the Z-scores for q SNPs.
ld	a q*q correlation matrix of the q SNPs.
n	a q*1 matrix containing the sample sizes used to get the Z-scores for the q SNPs. Can be a number if the q sample sizes are the same.

Value

conAH returns the p-value of testing AH.

Author(s)

Yangqing Deng and Wei Pan.

References

Deng, Y., Pan, W. (2018b). Significance Testing for Allelic Heterogeneity. Genetics. September 1, 2018 vol. 210 no. 1 25-32; https://doi.org/10.1534/genetics.118.301111.

---

conAHseq *conAHseq*

Description

Infer the number of causal SNPs with a sequential procedure using summary statistics and intersection-union tests.

Usage

```
conAHseq(Z,ld,n,k=6,alpha=0.05)
```

Arguments

Z	a q*1 matrix containing the Z-scores for q SNPs.
ld	a q*q correlation matrix of the q SNPs.
n	a q*1 matrix containing the sample sizes used to get the Z-scores for the q SNPs. Can be a number if the q sample sizes are the same.
k	a maximum number of causal SNPs.
alpha	the significance threshold.

Value

conAHseq returns the p-value of each step as well as the predicted number of causal SNPs.

Author(s)

Yangqing Deng and Wei Pan.

References

Deng, Y., Pan, W. (2018b). Significance Testing for Allelic Heterogeneity. Genetics. September 1, 2018 vol. 210 no. 1 25-32; https://doi.org/10.1534/genetics.118.301111.

`conAHseq2` *conAHseq2*

Description

A faster version of conAHseq, which considers less cases but gives the causal locations.

Usage

```
conAHseq2(Z,ld,n,k=6,alpha=0.05)
```

Arguments

`Z`	a q*1 matrix containing the Z-scores for q SNPs.
`ld`	a q*q correlation matrix of the q SNPs.
`n`	a q*1 matrix containing the sample sizes used to get the Z-scores for the q SNPs. Can be a number if the q sample sizes are the same.
`k`	a maximum number of causal SNPs.
`alpha`	the significance threshold.

Value

conAHseq2 returns the significant causal SNPs, the p-value of each step and the predicted number of causal SNPs.

Author(s)

Yangqing Deng and Wei Pan.

References

Deng, Y., Pan, W. (2018b). Significance Testing for Allelic Heterogeneity. Genetics. September 1, 2018 vol. 210 no. 1 25-32; https://doi.org/10.1534/genetics.118.301111.

JointSum *JointSum*

Description

Use summary statistics to build a joint linear model for one trait vs. L SNPs or one trait vs. L SNPs + (K-1) traits.

Usage

```
JointSum(B1,S1,B2=0,S2=0,N,XX=diag(1,nrow=1),YY0,adj_Y=1,lam=0)
```

Arguments

B1	a L*1 matrix containing marginal effects on the trait treated as response.
S1	a L*1 matrix containing standard errors for B1.
B2	a L*(K-1) matrix containing marginal effects on the (K-1) traits to adjust for. If K=1, do not specify this.
S2	a L*(K-1) matrix containing standard errors for B2. If K=1, do not specify this.
N	a L*K matrix containing sample sizes for each coefficients in B1, B2.
XX	a L*L estimated covariance matrix for the L SNPs.
YY0	a K*K estimated correlation matrix for the K traits.
adj_Y	whether traits should be adjusted for. If it is 0, adjust for SNPs only. Otherwise adjust for both SNPs and traits.
lam	a modifying parameter in [0,1). It is used only if adj_Y=1.

Value

beta	coefficient estimates (SNPs first).
cov	the covariance matrix for coefficients.
pvalue	p-values for coefficients.
sigma2	estimated mean squared error.

Author(s)

Yangqing Deng and Wei Pan.

References

Deng, Y., Pan, W. (2017). Conditional analysis of multiple quantitative traits based on marginal GWAS summary statistics. Genet Epidemiol. doi: 10.1002/gepi.22046.

Examples

```
#2 correlated SNPs, 2 traits
set.seed(13)
x1=rbinom(1000,1,0.3)
x2=c(x1[1:300],rbinom(1000-300,1,0.3))
y2=rnorm(1000)+x1
y1=rnorm(1000)+y2/2

#standardization
x1=x1-mean(x1)
x2=x2-mean(x2)
y1=y1-mean(y1)
y2=y2-mean(y2)

#summary statistics
a=summary(lm(y1~x1-1))$coefficients
b=summary(lm(y1~x2-1))$coefficients
c=summary(lm(y2~x1-1))$coefficients
d=summary(lm(y2~x2-1))$coefficients

B1=as.matrix(c(a[1],b[1]))
S1=as.matrix(c(a[2],b[2]))
B2=as.matrix(c(c[1],d[1]))
S2=as.matrix(c(c[2],d[2]))
N=matrix(1000,nrow=2,ncol=2)
XX=cov(cbind(x1,x2))
YY0=cor(cbind(y1,y2))

#model Y1 ~ X1 + X2
JointSum(B1,S1,B2,S2,N,XX,YY0,adj_Y=0)
#or
JointSum(B1,S1,N=N[,1],XX=XX,YY0=diag(1),adj_Y=0)

#model Y1 ~ X1 + X2 + Y2
JointSum(B1,S1,B2,S2,N,XX,YY0,adj_Y=1)

#may compare with joint models using individual level data
summary(lm(y1~x1+x2-1))
summary(lm(y1~x1+x2+y2-1))
```

Plei *Plei*

Description

Test pleiotropy with union-intersection tests and individual level data.

Usage

```
Plei(X,Y,pr=c(1:(nrow(YY0),nrow(YY0)*nrow(XX)+1),method="wald",ay=1,suby=c())
```

Arguments

X	a n*q matrix containing q SNPs for n subjects.
Y	a n*p matrix containing p traits for n subjects.
pr	a vector indicating which tests should be included in the union-intersection test. It is suggested to only use the default setting, which includes 1, 2, ..., p and p*q+1. 1, 2, ..., p correspond to testing whether the first SNP only influences one of the p traits. p*q+1 correspond to testing whether the first SNP influences none of the p traits.
method	the method to be used for each individual test. method = "lrt" uses the likelihood ratio test as described in Schaid et. al (2016); method = "wald" uses the Wald test with ordinary least square estimates and sandwich covariance matrices. It is recommended to use the Wald test when ay = 1.
ay	whether some of the traits should be adjusted for. ay = 1 means to use conditional analysis adjusting for traits (and possibly SNPs if q is bigger than 1); ay = 0 means not to adjust for any traits.
suby	a vector indicating which of the traits should NOT be adjusted for. It is only effective when ay = 1. If suby is not specified, all of the traits will be adjusted for in the conditional analysis (which is not recommended). If suby = c(1,3), the first and the third traits will not be adjusted for. (In the current version, suby only works for method = "wald"; if method = "lrt", suby is always considered as c())

Value

Plei returns a vector containing 2 p-values. The first p-value corresponds to testing whether the SNP of interest only influences one of the trait. The second p-value corresponds to testing whether the SNP does not have any effect. The p-value for the pleiotropy test is the maximum of these two values.

Author(s)

Yangqing Deng and Wei Pan.

References

Deng, Y., Pan, W. (2017). Testing Genetic Pleiotropy with GWAS Summary Statistics for Marginal and Conditional Analyses. Genetics. 2017 Dec; 207(4): 1285-1299. doi: 10.1534/genetics.117.300347.

Schaid, D. J., Tong, X., Larrabee, B., Kennedy, R. B., Poland, G. A., & Sinnwell, J.P. (2016). Statistical Methods for Testing Genetic Pleiotropy. Genetics 204(2): 483-497.

Examples

```
#2 correlated SNPs, 2 traits
set.seed(13)
x1=rbinom(1000,1,0.3)
x2=c(x1[1:300],rbinom(1000-300,1,0.3))
y2=rnorm(1000)+x1
y1=rnorm(1000)+y2/2
X=cbind(x1,x2)
Y=cbind(y1,y2)
```

```
#do not condition on traits (p-values for pleiotropy tests)
#y1 ~ x1 + x2; y2 ~ x1 + x2 (H0: none or only one of x1's coefficients is nonzero)
qq=Plei(X,Y,method="lrt",ay=0,suby=c())
max(qq)
qq=Plei(X,Y,method="wald",ay=0,suby=c())
max(qq)

#condition on both traits
#y1 ~ x1 + x2 + y2; y2 ~ x1 + x2 + y1
qq=Plei(X,Y,method="wald",ay=1,suby=c())
max(qq)

#condition on trait 2
#y1 ~ x1 + x2 + y2; y2 ~ x1 + x2
qq=Plei(X,Y,method="wald",ay=1,suby=c(1))
max(qq)
```

PleiSum *PleiSum*

Description

Test pleiotropy with union-intersection tests and summary statistics.

Usage

```
PleiSum(BM,SM,XX=diag(1),YY0,n,pr=c(1:(nrow(YY0),nrow(YY0)*nrow(XX)+1),method="wald",ay=1,suby=c
```

Arguments

BM	a p*q matrix containing marginal effect sizes (or Z-scores) from summary statistics. q is the number of SNPs, and p is the number of traits. The first SNP is to be tested, while the other (q-1) SNPs are treated as covariates to adjust for.
SM	a p*q matrix containing marginal standard errors corresponding to BM from summary statistics. If Z-scores are used for BM, all entries of SM should be set to 1.
XX	a q*q estimated covariance matrix for the q SNPs, the order of which should be consistent with BM.
YY0	a p*p estimated correlation matrix for the p traits, the order of which should be consistent with SM.
n	a number indicating the total sample size.
pr	a vector indicating which tests should be included in the union-intersection test. It is suggested to only use the default setting, which includes 1, 2, ..., p and p*q+1. 1, 2, ..., p correspond to testing whether the first SNP only influences one of the p traits. p*q+1 correspond to testing whether the first SNP influences none of the p traits.
method	the method to be used for each individual test. method = "lrt" uses the likelihood ratio test as described in Schaid et. al (2016); method = "wald" uses the Wald test with ordinary least square estimates and sandwich covariance matrices. It is recommended to use the Wald test when ay = 1.

ay — whether some of the traits should be adjusted for. ay = 1 means to use conditional analysis adjusting for traits (and possibly SNPs if q is bigger than 1); ay = 0 means not to adjust for any traits.

suby — a vector indicating which of the traits should NOT be adjusted for. It is only effective when ay = 1. If suby is not specified, all of the traits will be adjusted for in the conditional analysis (which is not recommended). If suby = c(1,3), the first and the third traits will not be adjusted for. (In the current version, suby only works for method = "wald"; if method = "lrt", suby is always considered as c())

Value

PleiSum returns a vector containing 2 p-values. The first p-value corresponds to testing whether the SNP does not have any effect. The second p-value corresponds to testing whether the SNP of interest only influences one of the trait. The p-value for the pleiotropy test is the maximum of these two values.

Author(s)

Yangqing Deng and Wei Pan.

References

Deng, Y., Pan, W. (2017). Testing Genetic Pleiotropy with GWAS Summary Statistics for Marginal and Conditional Analyses. Genetics. 2017 Dec; 207(4): 1285-1299. doi: 10.1534/genetics.117.300347.

Schaid, D. J., Tong, X., Larrabee, B., Kennedy, R. B., Poland, G. A., & Sinnwell, J.P. (2016). Statistical Methods for Testing Genetic Pleiotropy. Genetics 204(2): 483-497.

Examples

```
#2 correlated SNPs, 2 traits
set.seed(13)
x1=rbinom(1000,1,0.3)
x2=c(x1[1:300],rbinom(1000-300,1,0.3))
y2=rnorm(1000)+x1
y1=rnorm(1000)+y2/2

#standardization
x1=x1-mean(x1)
x2=x2-mean(x2)
y1=y1-mean(y1)
y2=y2-mean(y2)

#summary statistics
a=summary(lm(y1~x1-1))$coefficients
b=summary(lm(y1~x2-1))$coefficients
c=summary(lm(y2~x1-1))$coefficients
d=summary(lm(y2~x2-1))$coefficients

B1=as.matrix(c(a[1],b[1]))
S1=as.matrix(c(a[2],b[2]))
B2=as.matrix(c(c[1],d[1]))
S2=as.matrix(c(c[2],d[2]))
BM=t(cbind(B1,B2))
```

```
SM=t(cbind(S1,S2))
n=1000
XX=cov(cbind(x1,x2))
YY0=cor(cbind(y1,y2))

#do not condition on traits (p-values for pleiotropy tests)
#y1 ~ x1 + x2; y2 ~ x1 + x2 (H0: none or only one of x1's coefficients is nonzero)
qq=PleiSum(BM,SM,XX,YY0,n,method="lrt",ay=0,suby=c())
max(qq)
qq=PleiSum(BM,SM,XX,YY0,n,method="wald",ay=0,suby=c())
max(qq)

#condition on both traits
#y1 ~ x1 + x2 + y2; y2 ~ x1 + x2 + y1
qq=PleiSum(BM,SM,XX,YY0,n,method="wald",ay=1,suby=c())
max(qq)

#condition on trait 2
#y1 ~ x1 + x2 + y2; y2 ~ x1 + x2
qq=PleiSum(BM,SM,XX,YY0,n,method="wald",ay=1,suby=c(1))
max(qq)
```

SMI *SMI*

Description

Build a joint linear model for one trait vs. q SNPs using summary statistics and the MI-type approach.

Usage

```
SMI(BM,SM,N,Bref,mult=30)
```

Arguments

BM	a q*1 matrix containing marginal effects of the q SNPs on the trait.
SM	a q*1 matrix containing standard errors for BM.
N	a q*1 matrix containing sample sizes for each coefficient in BM.
Bref	a nref*q matrix. Reference data with nref subjects and the q SNPs.
mult	the number of imputations for the MI-type approach. When mult=0, do not use the MI-type approach.

Value

beta	coefficient estimates.
cov	the covariance matrix for coefficients.
chisq	the test statistic for the Wald test (jointly testing the q SNPs).
df	the degree of freedom for the test statistic.
pvalue	the p-value.

Author(s)

Yangqing Deng and Wei Pan.

References

Deng, Y., Pan, W. (2018a). Improved Use of Small Reference Panels for Conditional and Joint Analysis with GWAS Summary Statistics. Genetics. June 1, 2018 vol. 209 no. 2 401-408; https://doi.org/10.1534/genetics.118.300813.

Examples

```
#2 SNPs, 1 trait
set.seed(190)
x1=rbinom(1000,1,0.3)
x2=c(x1[1:400],rbinom(1000-400,1,0.3))
y1=rnorm(1000)
Bref0=cbind(x1,x2)

#standardization
x1=x1-mean(x1)
x2=x2-mean(x2)
y1=y1-mean(y1)

#summary statistics
a=summary(lm(y1~x1-1))$coefficients
b=summary(lm(y1~x2-1))$coefficients

BM=as.matrix(c(a[1],b[1]))
SM=as.matrix(c(a[2],b[2]))
N=matrix(1000,nrow=2,ncol=1)

#reference data
x1b=rbinom(500,1,0.3)
x2b=c(x1b[1:220],rbinom(500-220,1,0.3))
Bref=cbind(x1b,x2b)

#models using reference data
SMI(BM,SM,N,Bref,mult=1)
SMI(BM,SM,N,Bref,mult=30)

#model using original data
SMI(BM,SM,N,Bref0,mult=1)
```

Index