Package 'simGWAS'

August 22, 2019

Type Package

Title Fast Simulation of Large Case-Control GWAS Summary Statistics

Version 0.2.0-2

Date 2019-08-19

Author Mary Fortune [aut],
Chris Wallace [aut, cre],
Marcus Klarqvist [ctb]

Maintainer Chris Wallace <cew54@cam.ac.uk>

Description Simulating output from a case-control genome wide association study (GWAS) with a given causal model. Fortune and Wallace (2019) <doi:10.1093/bioinformatics/bty898>.

License GPL

LazyLoad yes

OS_type unix

Collate 'simGWAS.R' 'compute_gamma0.R' 'vbeta.R' 'zscore.R' 'extractsnps.R' 'make_dataset.R' 'make_GenoProbList.R' 'RcppExports.R'

LinkingTo Rcpp

Imports Rcpp, dplyr, combinat, corpcor, mvtnorm

RoxygenNote 6.1.1

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

URL https://github.com/chr1swallace/simGWAS

BugReports https://github.com/chr1swallace/simGWAS/issues

NeedsCompilation yes

Repository CRAN

Date/Publication 2019-08-22 08:20:02 UTC

R topics documented:

`compute_gamma0` *Compute gamma0, given haplotype frequencies*

Description

Compute the value of gamma_0 from gamma1,...,gamma_m

Usage

```
compute_gamma0(N0, N1, W, gamma.W, freq)
```

Arguments

`N0`	number of samples with Y=0
`N1`	number of samples with Y=1
`W`	The causal SNPs
`gamma.W`	The odds ratios of effect for each CV
`freq`	Frequencies of SNP appearances (computed using snphap)

Details

Note: assume we must compute the distibution of W in the controls

Value

The value of gamma0

Author(s)

Mary Fortune

compute_gamma0_PW *Compute gamma0 given distribution of W*

Description

Compute the value of gamma_0 from gamma1,...,gamma_m

Usage

```
compute_gamma0_PW(N0, N1, W, gamma.W, PWgY0)
```

Arguments

N0	number of samples with Y=0
N1	number of samples with Y=1
W	The causal SNPs
gamma.W	The odds ratios of effect for each CV
PWgY0	the distribution of causal snps in controls

Details

Note: assume we know the distibution of W in the controls

Value

The value of gamma0

Author(s)

Mary Fortune

est_statistic *estimate Z score at a single SNP*

Description

Wrapper function to run est_zscore for all snps in snps

Usage

```
est_statistic(N0, N1, snps, W, gamma.W, freq, GenoProbList)
```

Arguments

N0	The number of Y=0
N1	The number of Y=1
snps	The snps at which we wish to compute the expected Z Score
W	The true causal SNPs (these need not be in "snps")
gamma.W	The odds ratios of effect of the true causal SNPs (not including gamma0, the intercept term)
freq	Frequencies of SNP appearances (computed using snphap)
GenoProbList	An list of objects giving the probability of seeing each X,W genotype vector

Details

Assumes we have a list, GenoProbList, giving the GenoProb values for each X.

Value

The expected Z Score for SNP X, assuming the causal SNPs are W

Author(s)

Mary Fortune and Chris Wallace

---

est_zscore *estimate Z score at a single SNP*

Description

Estimates the expected Z Score for a single SNP

Usage

```
est_zscore(N0, N1, Ufactor, powerfactor, freq, GenoProbXW)
```

Arguments

N0	The number of Y=0
N1	The number of Y=1
Ufactor	The constant factor used to compute the expectation of U
powerfactor	The constant factor used to compute the expectation of the genotype of X to some power
freq	Frequencies of SNP appearances (computed using snphap)
GenoProbXW	An object giving the probability of seeing each X,W genotype vector

Details

Assumes the input CVs, and the relationship, gamma, between them and the trait of interest

Assumes we have already generated GenoProbXW for all X

Value

The expected Z Score for SNP X, assuming the causal SNPs are W

Author(s)

Mary Fortune and Chris Wallace

---

expected_vbeta *Estimate the expected variance of beta*

Description

Estimate the expected variance of beta. This is approximately expected(1/var(U)).

Usage

```
expected_vbeta(N0, N1, snps, W, gamma.W, freq,
  GenoProbList = make_GenoProbList(snps = snps, W = W, freq = freq))
```

Arguments

N0	The number of Y=0
N1	The number of Y=1
snps	The snps at which we wish to compute the expected Z Score
W	The true causal SNPs (these need not be in "snps")
gamma.W	The log odds ratios of effect of the true causal SNPs (not including gamma0, the intercept term)
freq	Haplotype frequencies as a data.frame, with column Probability indicating relative frequency in controls.
GenoProbList	An list of objects giving the probability of seeing each X,W genotype vector. This can be calculated within the function if no value supplied, or you can pass a pre-calculated version

Details

Assumes we have a list, GenoProbList, giving the GenoProb values for each X.

Value

The expected variance of beta for each SNP X, assuming the causal SNPs are W

Author(s)

Mary Fortune and Chris Wallace

Examples

```
freq=fake_freq(nhaps=100,nsnps=5) # fake haplotype frequency data
EVB=expected_vbeta(N0=1000,N1=2000,snps=paste0("s",1:5),
                   W="s1",gamma.W=log(1.5),freq=freq)
EVB # causal variant is SNP 1, with OR 1.5
```

---

expected_z_score *Compute vector of expected Z Scores*

Description

Compute vector of expected Z scores

Usage

```
expected_z_score(N0, N1, snps, W, gamma.W, freq,
  GenoProbList = make_GenoProbList(snps = snps, W = W, freq = freq))
```

Arguments

N0	The number of Y=0
N1	The number of Y=1
snps	The snps at which we wish to compute the expected Z Score
W	The true causal SNPs (these need not be in "snps")
gamma.W	The log odds ratios of effect of the true causal SNPs (not including gamma0, the intercept term)
freq	Haplotype frequencies as a data.frame, with column Probability indicating relative frequency in controls.
GenoProbList	An list of objects giving the probability of seeing each X,W genotype vector. This can be calculated within the function if no value supplied, or you can pass a pre-calculated version

Value

The expected Z Score for all snps in snps, assuming the causal SNPs are W

Author(s)

Mary Fortune and Chris Wallace

Examples

```
freq=fake_freq(nhaps=100,nsnps=5) # fake haplotype frequency data
EZ=expected_z_score(N0=1000,N1=2000,snps=paste0("s",1:5),
                    W="s1",gamma.W=log(1.5),freq=freq)
EZ # causal variant is SNP 1, with OR 1.5
```

`fake_freq` *fake haplotype frequencies*

Description

create a fake haplotype frequency dataset

Usage

```
fake_freq(nhaps = 100, nsnps = 10)
```

Arguments

`nhaps`	number of haplotypes
`nsnps`	number of snps

Details

no attempt is made at biological realism, this is purely for testing code

Value

data.frame of 1,2, nhaps x nsamples + frequency column

Author(s)

Chris Wallace

Examples

```
freq=fake_freq(nhaps=100,nsnps=5)
dim(freq)
head(freq)
```

make_GenoProbList *make_GenoProbList*

Description

compute a list, GenoProbList, giving the GenoProb values for each X.

Usage

```
make_GenoProbList(snps, W, freq)
```

Arguments

snps	Index of SNPs at which we wish to compute the expected Z Score
W	Index of true causal SNPs (these need not be in "snps")
freq	Frequencies of SNP appearances (computed using snphap)

Value

The the GenoProb values for each X

Author(s)

Mary Fortune and Chris Wallace

Examples

```
freq=fake_freq() # fake haplotype frequency data
problist=make_GenoProbList(1:2,W=1,freq)
```

simulated_vbeta *Compute a simulated var(beta)*

Description

Simulate var(beta)

Usage

```
simulated_vbeta(N0, N1, snps, W, gamma.W, freq,
  GenoProbList = make_GenoProbList(snps = snps, W = W, freq = freq),
  nrep = 1)
```

Arguments

N0	The number of Y=0
N1	The number of Y=1
snps	The snps at which we wish to compute the expected Z Score
W	The true causal SNPs (these need not be in "snps")
gamma.W	The log odds ratios of effect of the true causal SNPs (not including gamma0, the intercept term)
freq	Haplotype frequencies as a data.frame, with column Probability indicating relative frequency in controls.
GenoProbList	An list of objects giving the probability of seeing each X,W genotype vector. This can be calculated within the function if no value supplied, or you can pass a pre-calculated version
nrep	Number of replicates (simulated vectors of Z scores) under this scenario. Default=1

Details

Assumes we have a list, GenoProbList, giving the GenoProb values for each X.

Value

A simulated variance of beta for each SNP X, assuming the causal SNPs are W

Author(s)

Mary Fortune and Chris Wallace

Examples

```
freq=fake_freq(nhaps=100,nsnps=5) # fake haplotype frequency data
VB=simulated_vbeta(N0=1000,N1=2000,snps=paste0("s",1:5),
                   W="s1",gamma.W=log(1.5),freq=freq)
VB # causal variant is SNP 1, with OR 1.5
```

---

simulated_z_null *Compute a NULL simulated Z Score*

Description

Compute matrix of simulated Z scores about expected values of 0 - ie under a null of no association at any SNP

Usage

```
simulated_z_null(snps, freq, nrep = 1)
```

Arguments

snps	The snps at which we wish to compute the expected Z Score
freq	Haplotype frequencies as a data.frame, with column Probability indicating relative frequency in controls.
nrep	Number of replicates (simulated vectors of Z scores) under this scenario. Default=1

Author(s)

Mary Fortune and Chris Wallace

Examples

```
freq=fake_freq(nhaps=100,nsnps=5) # fake haplotype frequency data
    Z=simulated_z_null(snps=paste0("s",1:5),freq=freq,nrep=3)
Z # no causal variants
```

---

simulated_z_score *Compute a simulated Z Score*

Description

Compute matrix of simulated Z scores

Usage

```
simulated_z_score(N0, N1, snps, W, gamma.W, freq,
  GenoProbList = make_GenoProbList(snps = snps, W = W, freq = freq),
  nrep = 1)
```

Arguments

N0	The number of Y=0
N1	The number of Y=1
snps	The snps at which we wish to compute the expected Z Score
W	The true causal SNPs (these need not be in "snps")
gamma.W	The log odds ratios of effect of the true causal SNPs (not including gamma0, the intercept term)
freq	Haplotype frequencies as a data.frame, with column Probability indicating relative frequency in controls.
GenoProbList	An list of objects giving the probability of seeing each X,W genotype vector. This can be calculated within the function if no value supplied, or you can pass a pre-calculated version
nrep	Number of replicates (simulated vectors of Z scores) under this scenario. Default=1

Author(s)

Mary Fortune and Chris Wallace

Examples

```
freq=fake_freq(nhaps=100,nsnps=5) # fake haplotype frequency data
    Z=simulated_z_score(N0=1000,N1=2000,snps=paste0("s",1:5),
                        W="s1",gamma.W=log(1.5),freq=freq,nrep=3)
    Z # causal variant is SNP 1, with OR 1.5
```

---

which_genotypes *maps haplotype pairs to genotypes*

Description

computes a matrix telling us which haplotype pairs correspond to which genotypes

Usage

```
which_genotypes(nsnps)
```

Arguments

nsnps	The number of SNPs

Details

Internal function

Author(s)

Mary Fortune

Index