Package 'powerGWASinteraction'

July 23, 2025

Version 1.1.3

Date 2015-07-02

Title Power Calculations for GxE and GxG Interactions for GWAS

Author Charles Kooperberg <clk@fredhutch.org> and Li Hsu <lih@fredhutch.org>

Maintainer Charles Kooperberg <clk@fredhutch.org>

Depends R (>= 2.10), mvtnorm, pwr

Imports stats

Description Analytical power calculations for GxE and GxG interactions for case-control studies of candidate genes and genome-wide association studies (GWAS). This includes power calculation for four two-step screening and testing procedures. It can also calculate power for GxE and GxG without any screening.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2015-07-03 01:34:42

Contents

`powerGE` *Power for GxE interactions in genetic association studies*

Description

This routine carries out (analytical, approximate) power calculations for identifying Gene-Environment interactions in Genome Wide Association Studies

Usage

```
powerGE(n, power, model, caco, alpha, alpha1, maintain.alpha)
```

Arguments

n	Sample size: combined number of cases and controls. Note: exactly one of n and power should be specified.
power	Power: targeted power. Note: exactly one of n and power should be specified.
model	List specifying the genetic model. This list contains the following objects: • prev Prevalence of the outcome in the *population*. Note that for case-only and empirical Bayes estimators to be valid, the prevalence needs to be low. • pGene Probability that a *binary* SNP is 1 (i.e. not the minor allele frequency for a three level SNP). • pEnv Frequency of the binary environmental variable. • orGE Odds ratio between the binary SNP and binary environmental variable. • beta.LOR Vector of length three with the odds ratios of the genetic, environmental, and GxE interaction effect, respectively. • nSNP Number of SNPs (genes) being tested.
caco	Fraction of the sample that are cases (default = 0.5).
alpha	Overall (family-wise) Type 1 error (default = 0.05).
alpha1	Significance level at which testing during the first stage (screening) takes place. If alpha1 = 1, there is no screening.
maintain.alpha	Some combinations of screening and GxE testing methods do not maintain the proper Type 1 error. Default is True: combinations that do not maintain the Type 1 error are not computed. If maintain.alpha is False all combinations are computed.

Details

The routine computes power for a variety of two-stage procedures. Five different screening procedures are used:

- **No screening** All SNPs are tested for interaction
- **Marginal screening** Only SNPs that are marginally significant at level alpha1 are screened for interaction. See Kooperberg and LeBlanc (2010).
- **Correlation screening** Only SNPs that are, combined over all cases and controls, associated with the environmental variable at level alpha 1 are screened for interaction. See Murcray et al. (2012).
- **Cocktail screening** SNPs are screened on the most significant of marginal and correlation screening. See Hsu et al. (2012).
- **Chi-square screening** SNPs are screened using a chi-square combination of correlation and marginal screening. See Gauderman et al. (2013).

After screening, the SNPs that pass the screen can be tested using

- **Case-control** The standard case-control estimator.
- **Case-only** The case-only estimator.
- **Empirical Bayes** The empirical Bayes estimator of Mukherjee and Chatterjee (2010).

If screening took place using the correlation or chi-square screening, the Type 1 error won't be maintained if the final GxE testing is carried out using either the case-only or empirical Bayes estimator. See Dai et al. (2012). The cocktail screening maintains the Type 1 family wise error rate, since only those SNPs that pass on to the second stage using marginal screening will use the case-only or empirical Bayes estimator, the SNPs that pass on to the second stage using correlation screening will always use the case-control estimator.

When SNP and environment are correlated in the population (i.e. `model$orGE` does not equal 1) the case-only estimator does not maintain the Type 1 error. The empirical Bayes estimator may also have a moderately inflated Type 1 error. When the disease is common either the case-only estimator or the empirical Bayes estimator also may not estimate the GxE interaction.

Power calculations are described in Kooperberg, Dai, and Hsu (2014). Briefly, for a given genetic model we compute the expected p-values for all screening statistics. We then use a normal approximation to compute the probability that this SNP passes the screening (e.g., if `alpha1` equaled this expected p-value this probability would be exactly 0.5), and combine this with power calculations for the second stage of GxE testing.

Value

A list with three components.

`power`	A 5x3 matrix with estimated power for all testing approaches, only if `n` was specified.
`samplesize`	A 5x3 matrix with required sample sizes for all testing approaches, only if `power` was specified.
`expected.p`	A 5x3 matrix with the expected p value for the SNP to pass screening. This p-value depends on the sample size, but not on the second stage testing.
`prob.select`	A 5x3 matrix with the probability that the interacting SNP would pass the screening stage. This probability depends on the sample size, but not on the second stage testing.

Author(s)

Li Hsu <lih@fredhutch.org> and Charles Kooperberg <clk@fredhutch.org>.

References

Dai JY, Kooperberg C, LeBlanc M, Prentice RL (2012). Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika*, **99**, 929-944.

Gauderman WJ, Zhang P, Morrison JL, Lewinger JP (2013). Finding novel genes by testing GxE interactions in a genome-wide association study. *Genetic Epidemiology*, **37**, 603-613.

Hsu L, Jiao S, Dai JY, Hutter C, Peters U, Kooperberg C (2012). Powerful cocktail methods for detecting genome-wide gene-environment interaction. *Genetic Epidemiology*, **36**, 183-194.

Kooperberg C, Dai, JY, Hsu L (2014). Two-stage procedures for the identification of gene x environment and gene x gene interactions in genome-wide association studies. *To appear*.

Kooperberg C, LeBlanc ML (2008). Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genetic Epidemiology*, **32**, 255-263.

Mukherjee B, Chatterjee N (2008). Exploiting gene-environment inde- pendence for analysis of case-control studies: an empirical Bayes-type shrinkage estimator to trade-off between bias and efficiency *Biometrics*, **64**, 685-694.

Murcray CE, Lewinger JP, Gauderman WJ (2009). Gene-environment interaction in genome-wide association studies. *American Journalk of Epidemiology*, **169**, 219-226.

See Also

powerGG

Examples

```
mod1 <- list(prev=0.01,pGene=0.2,pEnv=0.2,beta.LOR=log(c(1.0,1.2,1.4)),orGE=1.2,nSNP=10^6)
results <- powerGE(n=20000, model=mod1,alpha1=.01)
print(results)

mod2 <- list(prev=0.01,pGene=0.2,pEnv=0.2,beta.LOR=log(c(1.0,1.0,1.4)),orGE=1,nSNP=10^6)
results <- powerGE(power=0.8, model=mod2,alpha1=.01)
print(results)
```

`powerGG` *Power for GxG interactions in genetic association studies*

Description

This routine carries out (analytical, approximate) power calculations for identifying Gene-Gene interactions in Genome Wide Association Studies

Usage

```
powerGG(n, power, model, caco, alpha, alpha1)
```

Arguments

`n`	Sample size: combined number of cases and controls. Note: exactly one of `n` and `power` should be specified.
`power`	Power: targeted power. Note: exactly one of `n` and `power` should be specified.
`model`	List specifying the genetic model. This list contains the following objects:

- `prev` Prevalence of the outcome in the *population*. Note that for case-only and empirical Bayes estimators to be valid, the prevalence needs to be low.
- `pGene1` Probability that the first *binary* SNP is 1 (i.e. not the minor allele frequency for a three level SNP).

- pGene2 Probability that the first *binary* SNP is 1 (i.e. not the minor allele frequency for a three level SNP).
- beta.LOR Vector of length three with the odds ratios of the first genetic, second genetic, and GxG interaction effect, respectively.
- nSNP Number of SNPs (genes) being tested.

caco — Fraction of the sample that are cases (default = 0.5).

alpha — Overall (family-wise) Type 1 error (default = 0.05).

alpha1 — Significance level at which testing during the first stage (screening) takes place. If alpha1 = 1, there is no screening.

Details

The routine computes power calculations for a two-stage procedure with marginal screening followed by either case-control or case-only testing.

Value

A data frame consisting of two numbers: the power for the case-control and case-only approaches if n is specified or the required combined sample size for the case-control and case-only approaches if power is specified.

Author(s)

Charles Kooperberg, <clk@fredhutch.org>

References

Kooperberg C, LeBlanc M (2008). Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genetic Epidemiology*, **32**, 255-263.

See Also

powerGG

Examples

```
mod1 <- list(prev=0.05, pGene1=0.3, pGene2=0.3, beta.LOR=c(0,0,.6),nSNP=500000)
powerGG(n=10000,mod=mod1,caco=0.5,alpha=.05,alpha1=.001)
powerGG(power=0.8,mod=mod1,caco=0.5,alpha=.05,alpha1=.001)
```

`powerGWASinteraction` *Depreciated*

Description

This function is depreciated and has been replaced by `powerGG` and `powerGE`

Usage

```
powerGWASinteraction()
```

Value

An error message is printed

Author(s)

Charles Kooperberg, <clk@fredhutch.org>

See Also

powerGG,powerGE

Examples

```
powerGWASinteraction()
```

Index