LPG Packages: Implement a four-group probabilistic approach to leveraging pleiotropy in genome-wide association studies

Yi Yang[1], Mingwei Dai[2,4] Jian Huang[3] Xinyi Lin[5] Can Yang[4] Jin Liu[5] Min Chen[1]**

2017-10-13

[1]School of Statistics and Management, The Shanghai University of Finance and Economics, Shanghai
[2]Institute for Information and System Sciences, Xian Jiaotong University, Xian
[3]Department of Applied Mathematics, Hong Kong Polytechnics University
[4]Department of Mathematics, Hong Kong University of Science and Technology
[5]Centre for Quantitative Medicine, Duke-NUS Medical School

Overview

The vignette provides an introduction to R package `LPG` . The `LPG` package provdes functions to analyze pleiotropic effects among GWAS data sets and than make inference about pleiotropy.

install the packages

```
install.packages("devtools")
library(devtools)
install_github("Shufeyangyi2015310117/LPG")
```

load the package

```
library(LPG)
```

Data introduction

There are two classes of data that LPG can analyze. The first is ordinary matrix often used in usual linear regression, The other is plink file often used in GWAS study.

Fitting the LPG with matrix data

There is simulation data in the package, we can load it using the command

```
data(simulation)
```

we can get two n-by-p deisgn matrix X, X2, two n-by-1 phenotype y, y2 and two n-by -q covariates matrix z, z2.

We can fit different model with following commands. In addition, we can set the opts to satisfiy our demand. Fit Ftp for separate analysis with quantitative phenotype

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5)
fit<-Lpg(X, y, opts = opts)
```

Fit Ftp for separate analysis with binary phenotype

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5)
fit<-Lpg(X, y, family = "binomial", opts = opts)
```

Fit Ftp for joint analysis with quantitative phenotype

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5, constraintalpha = 0)
fit<-Lpg(X, y, x2 = X2, y2 = y2, opts = opts)
```

Fit Ftp for joint analysis with binary phenotype

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5, constraintalpha = 0)
fit<-Lpg(X, y, x2 = X2, y2 = y2, family = "binomial", opts = opts)
```

If we want to add covariates in Ftp, we can do as the following commmand.
Separate analysis

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5)
fit<-Lpg(X, y, z=z, family = "binomial", opts = opts)
```

Joint analysis

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5, constraintalpha = 0)
fit<-Lpg(X, y, z = z, x2 = X2, y2 = y2, z2 = z2, family = "binomial", opts = opts)
```

If we want to fit Ftp with alpha controlled. we set the logic paramter constraintalpha equal to 1.

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5, constraintalpha = 1)
fit<-Lpg(X, y, z = z, x2 = X2, y2 = y2, z2 = z2, family = "binomial", opts = opts)
```

The result of fit is a list contains model parameters and varaitional parameters. In general, it has following components.

```
str(fit)
```

```
## List of 7
##  $ vardist_gamma     : num [1:2000, 1:2] 0.000476 0.000724 0.000485 0.001242 0.000457 ...
##  $ vardist_mu        : num [1:2000, 1:2] -0.0713 -0.1894 -0.0619 -0.2769 -0.0199 ...
##  $ vardist_sigma2beta: num [1:2000, 1:2] 0.0362 0.0365 0.0399 0.0383 0.0373 ...
##  $ sigma2beta        : num [1:2, 1] 0.73 1.97
##  $ alpha             : num [1, 1:4] 9.96e-01 2.06e-08 1.68e-03 2.03e-03
##  $ Lq                : num [1, 1:34] -1380 -1340 -1331 -1324 -1319 ...
##  $ u                 : num [1:2, 1] -0.092 -0.00362
```

Fitting the LPG with plink file

If we have two plink file BDqc37 and BDqc36 placed in the directory "D:/realdata/WTCCC_all", we can fit Ftp with the following commmand.
Separate analysis for quantitative phenotype

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5)
fit <- Lpg.Plink("D:/realdata/WTCCC_all/BDqc37", opts = opts)
```

Separate analysis for binary phenotype

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5)
fit <- Lpg.Plink("D:/realdata/WTCCC_all/BDqc37", family = "binomial", opts = opts)
```

Joint analysis for quantitative phenotype

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5, constraintalpha = 0)
fit <- Lpg.Plink("D:/realdata/WTCCC_all/BDqc37",file2 = "D:/realdata/WTCCC_all/BDqc36", opts = opts)
```

Joint analysis for binary phenotype

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5, constraintalpha = 0)
fit <- Lpg.Plink("D:/realdata/WTCCC_all/BDqc37", file2 =
                   "D:/realdata/WTCCC_all/BDqc36", family = "binomial", opts = opts)
```

Separate analysis for binary phenotype with covariates

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5)
fit <- Lpg.Plink("D:/realdata/WTCCC_all/BDqc37", z, family = "binomial", opts = opts)
```

Joint analysis for binary phenotype with covariates

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5, constraintalpha = 0)
fit <- Lpg.Plink("D:/realdata/WTCCC_all/BDqc37", z = z, file2 =
                   "D:/realdata/WTCCC_all/BDqc36", z2 = z2, family = "binomial", opts = opts)
```

Get matrix data from plink file

If we we want to read a plink file to check it's content. We can read the plink file as following.

```
data <- Read.Plink("D:/realdata/WTCCC_all/BDqc37")
```

The object data is a list consisting of two conponents, genotype matrix "X" and phenotype "Y".

```
str(data)
```

```
## List of 2
##  $ X: num [1:4707, 1:446] 0 0 0 0 0 0 2 1 0 0 ...
##  $ Y: num [1:4707, 1] 0.705 0.554 0.696 0.694 0.584 ...
```

Statistical inference and Pleiotropy test

Association mapping

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5, constraintalpha = 0)
fit<-Lpg(X, y, x2 = X2, y2 = y2, family = "binomial", opts = opts)
```

Risk SNPs

```
riskmat<-assoc(fit, FDR = 0.2, fdrControl="global")
```

```
## Info: Association mapping based on the global FDR control at level 0.2.
```

struct of risk SNP matrix

```
str(riskmat)
```

```
##  num [1:2000, 1:2] 0 0 0 0 0 0 0 0 0 0 ...
```

Estimate under null hypothesis and alternative hypothesis

```
opts = list(max_iter = 1e5, dispF = 1, display_gap = 20, epsStopLogLik = 1e-5, constraintalpha = 1)
fit0<-Lpg(X, y, x2 = X2, y2 = y2, family = "binomial", opts = opts)
```

Pleiotropy test

```
out<-Pleiotropy.test(fit0,fit)
```

a list consists of LRT statistics and p-value

```
str(out)
```

```
## List of 2
##  $ LRT   : num 17.2
##  $ pvalue: num 3.4e-05
```

Reference

LPG: a four-group probabilistic approach to leveraging pleiotropy in genome-wide association studies