Composite Kernel Machine Regression based on Likelihood Ratio Test with Application for Combined Genetic and Gene-environment Interaction Effect

Ni Zhao *Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA*
Haoyu Zhang *Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA*
Jennifer J. Clark *Food and Drug Administration, Baltimore, MD, USA*
Arnab Maity *Department of Statistics, North Carolina State University, Raleigh, NC,USA*
Michael C. Wu *Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA*

Most common human diseases are a result from the combined effect of genes, the environmental factors and their interactions such that including gene-environment (GE) interactions can improve power in gene mapping studies. The standard strategy is to test the SNPs, one-by-one, using a regression model that includes both the SNP effect and the GE interaction. However, the SNP-by-SNP approach has serious limitations, such as the inability to model epistatic SNP effects, biased estimation and reduced power. Thus, in this paper, we develop a kernel machine regression framework to model the overall genetic effect of a SNP-set, considering the possible GE interaction. Specifically, we use a composite kernel to specify the overall genetic effect via a nonparametric function and we model additional covariates parametrically within the regression framework. The composite kernel is constructed as a weighted average of two kernels, one corresponding to the genetic main effect and one corresponding to the GE interaction effect. We propose a likelihood ratio test (LRT) and a restricted likelihood ratio test (RLRT) for statistical significance. We derive a Monte Carlo approach for the finite sample distributions of LRT and RLRT statistics. Extensive simulations and real data analysis show that our proposed method has correct type I error and can have higher power than score-based approaches under many situations.

Keywords: gene-environment interactions, kernel machine testing, likelihood ratio test, multiple variance compo- nents, spectral decomposition, unidentifiable conditions

---

Overview

This vegnette provides an introduction to the 'CKLRT' package. To load the package, users need to install package from CRAN and CKLRT from github. The package can be loaded with the following command:

```
#install.packages("devtools")
library(devtools)
#install_github("andrewhaoyu/CKLRT")
library(CKLRT)
```

Example

In this vegnette, we will decomstrate the methods with a simple example. 1. X present other covariates we want to adjust. 2. E represents the environment variable. 3. G is is the genotype matrix with two SNPs inside 4. y is the simulated outcomes.

```r
library(mgcv); library(MASS); library(nlme);
```

```
## Loading required package: nlme

## This is mgcv 1.8-18. For overview type 'help("mgcv-package")'.
```

```r
library(compiler);library(Rcpp);library(RcppEigen)
library(CKLRT)
set.seed(6)
n = 200 # the number of observations
X = rnorm(n) # the other covariates
p = 2 # two snp in a gene will be simulated
G = runif(n*p)< 0.5
G = G + runif(n*p) < 0.5
G = matrix(G, n,p) #genetic matrix
E = (runif(n) < 0.5)^2 #enviroment effect
y = rnorm(n) + G[,1] * 0.3 #observations
#apply the likelihood ratio test
omniLRT_fast(y, X =  cbind(X, E),K1 = G %*% t(G),K2 = (G*E) %*% t(G * E))
```

```
## $p.dir
## [1] 0.0248
##
## $p.aud
## [1] 0.02538226
##
## $LR
## [1] 3.408
```

```r
#apply the restricted likelihood ratio test
omniRLRT_fast(y, X =  cbind(X, E),K1 = G %*% t(G),K2 = (G*E) %*% t(G * E))
```

```
## $p.dir
## [1] 0.0225
##
## $p.aud
## [1] 0.0216479
##
## $LR
## [1] 3.701362
```

The results of the function contain three elements: 1. p.dir is the p-value of likelihood ratio test based on emprical distrition. 2. p.aud is the p-value by approximating the null distribution as a mixture of a point mass at zero with probability b and weighted chi square distribution with d degrees of freedom with probality of 1-b. 3. LR is the likelihood ratio test statistics.

References

1. N. Zhao, H. Zhang, J. Clark, A. Maity, M. Wu. Composite Kernel Machine Regression based on Likelihood Ratio Test with Application for Combined Genetic and Gene-environment Interaction Effect (Submitted)