Package ‘gcatest’

May 1, 2024

Title Genotype Conditional Association TEST
Version 2.4.0
Encoding UTF-8
LazyData true
Description GCAT is an association test for genome wide association studies that controls for population structure under a general class of trait models. This test conditions on the trait, which makes it immune to confounding by unmodeled environmental factors. Population structure is modeled via logistic factors, which are estimated using the `lfa` package.

Imports methods, lfa
Depends R (>= 4.0)
Suggests knitr, ggplot2, testthat, BEDMatrix, genio
VignetteBuilder knitr
License GPL (>= 3)
biocViews SNP, DimensionReduction, PrincipalComponent, GenomeWideAssociation

BugReports https://github.com/StoreyLab/gcatest/issues
URL https://github.com/StoreyLab/gcatest
Roxygen list(markdown = TRUE)
RoxygenNote 7.2.3
git_url https://git.bioconductor.org/packages/gcatest
git_branch RELEASE_3_19
git_last_commit 9ae4c6a
git_last_commit_date 2024-04-30
Repository Bioconductor 3.19
Date/Publication 2024-05-01
Author Wei Hao [aut],
Minsun Song [aut],
Alejandro Ochoa [aut, cre] (<https://orcid.org/0000-0003-4928-3403>),
John D. Storey [aut] (<https://orcid.org/0000-0001-5992-402X>)
Maintainer Alejandro Ochoa <alejandro.ochoa@duke.edu>
delta_deviance_lf

Calculate delta deviance of logistic null/alternative models

Description

This function fits, at each locus of a given genotype matrix, two logistic models, and under the assumption that the models are nested, calculates the delta deviance between the two. This general function is intended for testing models in a broad setting; for the specific problem of genetic association, the interface in \texttt{gcat()} and \texttt{gcat.stat()} are more user-friendly.

Usage

\begin{verbatim}
delta_deviance_lf(X, LF0, LF1)
\end{verbatim}

Arguments

\begin{itemize}
  \item \textbf{X} \hspace{1cm} A matrix of SNP genotypes, i.e. an integer matrix of 0’s, 1’s, 2’s and NAs. BEDMatrix is supported. Sparse matrices of class Matrix are not supported (yet).
  \item \textbf{LF0} \hspace{1cm} Logistic factors for null model.
  \item \textbf{LF1} \hspace{1cm} Logistic factors for alternative model.
\end{itemize}

Value

The vector of delta deviance values, one per locus of \texttt{X}.

Examples

\begin{verbatim}
library(lfa)

# make example data smaller so example is fast  
# goes from 1000 to 100 individuals
indexes <- sample.int(ncol(sim_geno), 100)

sim_geno <- sim_geno[, indexes]
sim_trait <- sim_trait[indexes]

# now run LFA and get delta deviances for trait assoc
# (recapitulating \texttt{gcat.stat} in this case)
LF <- lfa(sim_geno, 3)
LF0 <- LF # structure is null
\end{verbatim}
gcat <- cbind(LF, sim_trait)  # trait is alt
devdiff_assoc <- delta_deviance_lf(sim_geno, LF0, LF1)

# can instead do delta deviances for structure only
LF0 <- cbind(rep.int(1, ncol(sim_geno)))  # intercept only is null
LF1 <- LF  # structure is alt, no trait
devdiff_struc <- delta_deviance_lf(sim_geno, LF0, LF1)

---

gcat  
**Genotype Conditional Association TEST**

---

**Description**

Performs the GCAT association test between SNPs and trait, returning p-values.

**Usage**

```
 gcat(X, LF, trait, adjustment = NULL)
 gcatest(X, LF, trait, adjustment = NULL)
 gcat.stat(X, LF, trait, adjustment = NULL)
```

**Arguments**

- **X**: A matrix of SNP genotypes, i.e. an integer matrix of 0's, 1's, 2's and NAs. BEDMatrix is supported. Sparse matrices of class Matrix are not supported (yet).
- **LF**: matrix of logistic factors from `lfa::lfa()`
- **trait**: vector
- **adjustment**: matrix of adjustment variables

**Value**

vector of p-values

**Functions**

- `gcatest()`: Alias of gcat
- `gcat.stat()`: returns the association statistics instead of the p-value.

**References**

Examples

library(lfa)

# make example data smaller so example is fast
# goes from 1000 to 100 individuals
indexes <- sample.int(ncol(sim_geno), 100)
sim_geno <- sim_geno[, indexes]
sim_trait <- sim_trait[indexes]

# now run LFA and GCA Test
LF <- lfa(sim_geno, 3)
gcat_p <- gcat(sim_geno, LF, sim_trait)
gcat_stat <- gcat.stat(sim_geno, LF, sim_trait)

---

**sim_geno**  
Simulated data from PSD model

---

**Description**

10,000 SNPs, 1,000 individuals, first five SNPs are associated.

**Usage**

sim_geno

**Format**

a matrix of 0’s, 1’s and 2’s for the genotypes

**Value**

simulated genotype matrix

---

**sim_trait**  
Simulated data from PSD model

---

**Description**

10,000 SNPs, 1,000 individuals, first five SNPs are associated.

**Usage**

sim_trait

**Format**

a vector of traits
Value
  simulated traits
Index

delta_deviance_lf, 2

gcat, 3
gcat(), 2
gcat.stat(), 2
gcatest (gcat), 3
lfal:lfal(), 3

sim_genoe, 4
sim_trait, 4