1 Introduction

The \texttt{gcatest} package provides an implementation of the Genotype Conditional Association Test (GCAT) \cite{cite}. GCAT is a test for genetic association that is powered by Logistic Factor Analysis (LFA) \cite{cite}. LFA is a method of modeling population structure in a genome wide association study. GCAT performs a test for association between each SNP and a trait (either quantitative or binary). We have shown that GCAT is robust to confounding from population structure.

2 Sample usage

We include a sample dataset with the package. \texttt{sim\_geno} is a simulated genotype matrix. \texttt{sim\_trait} is a simulated trait. There are 10,000 SNPs and 1,000 individuals. The first five SNPs are associated with the trait. This simulations were done under the Pritchard-Stephens-Donnelly model with $K = 3$, with Dirichlet parameter $\alpha = 0.1$ and variance allotment in the trait corresponding to 5% genetic, 5% environmental, and 90% noise. This dataset is simulated under identical parameters as the PSD simulation in Figure 2 of the paper \cite{cite}, except that we have adjusted the size of the simulation to be appropriate for a small demo.

```r
library(lfa)
library(gcatest)
dim(sim\_geno)
## [1] 10000 1000
length(sim\_trait)
## [1] 1000
```

2.1 \texttt{gcat}

The first step of \texttt{gcat} is to estimate the logistic factors:

```r
LF <- lfa(sim\_geno, 3)
dim(LF)
## [1] 1000 3
```

Then, we call the \texttt{gcat} function, which returns a vector of p-values:

```r
```
gcat_p <- gcat(sim_geno, LF, sim_trait)

We can look at the p-values for the associated SNPs:

gcat_p[1:5]
## [1] 5.873675e-01 1.570672e-01 1.240494e-03 4.799294e-01 8.607748e-11

And also plot the histogram of the unassociated SNPs:

library(ggplot2)
dat <- data.frame(p = gcat_p[6:10000])
ggplot(dat, aes(p, after_stat(density))) + geom_histogram(binwidth=1/20) + theme_bw()

3 Data Input

The genio package provides the function read_plink for parsing PLINK binary genotypes (extension: .bed) into an R object of the format needed for the gcat function. A BEDMatrix object (from the eponymous function and package) is also supported, and can result in reduced memory usage (at a small runtime penalty).

References
