Evaluating Statistical Significance in Two-Stage Genomewide Association Studies

1. General description

The program *gas2.f* was written in standard FORTRAN-77. No external subroutines or functions are used. All the data are stored in the array A, whose dimension may be modified according to the memory constraint of your machine and the size of your data. Most machines will not allow a very large number of markers. To alleviate the problem of memory constraint, it is advised that the program be run on each chromosome separately.

2. Input

The program requires two separate groups of input: the data input and the control parameters input. Each record in the data file provides the trait value, followed by the genotype scores for the markers. Under the dominant model, the genotype score is 1 if the subject has at least one minor allele and 0 otherwise; under the recessive model, the genotype score is 1 if the subject has two minor alleles and 0 otherwise; under the additive model, the genotype score is the number of minor alleles. The genotype score is set to 9 if its value is unknown. The records for the subjects that are genotyped in stage 1 should appear first in the data file. The data should be in free format, i.e., the data items are separated by spaces or commas.

The control parameters described below are given in a file called *gas2.dat*.

Parameter	Type
problem title	character
file name of data input	character
file name of program output	character
total number of subjects	integer
total number of markers	integer
number of subjects in stage 1	integer
critical value for stage 1	real
number of Monte Carlo samples	integer

3. Output

Computational results are written to the output file specified by the user. For each marker, the output shows the observed value of the chi-squared test statistic, the (unadjusted) p-value, the p-values adjusted for multiple testing. Two adjusted p-values are presented: one is based on Holm's step-down procedure (a slightly less conservative version of the Bonferroni correction), and one is based on the Monte Carlo procedure described in Lin (2006). Only the markers whose Monte Carlo adjusted p-values are less than 0.5 are displayed.

4. Test data set

test.dat is a test data set simulated from the HapMap data. The control parameters are given in *gas2.dat*, and the output file is *test.out*.