Last update: 7/18/04

MMAP interaction analysis

MMAP implements both fixed and random effects interaction models. Interactions can be fit as a single model and GxE can be run genome wide. We have also implemented analysis of variance-heterogeneity QTLs, called vQTLs, which will be fully documented pending submission of a manuscript.

Fixed effect interactions

--covariates <covariates>

List of covariates included in the model.

--interactions <interaction terms>

Interactions are coded using the '*' delimiter. For example, AGE*SEX models the interaction between the covariates AGE and SEX and AGE*SEX*BMI the 3-way interaction between AGE,SEX, and BMI. Interaction covariates do not need to be modeled as main effects nor 2-way interactions in the case of the 3-way interaction, though it is standard to include the combinatorial possibilities.

GxE interactions

--gxe_interaction <covariate>

Run a Gx<covariate> GWA, where G is determined from the model options and genotype file. This option currently requires that <covariate> is also listed in the –covariate option. The output will include the GxE beta, standard error, p-value and covariance between the G beta and E beta. This option requires the option **--binary_covariate_filename <file>** where <file> is an MxS binary genotype file.

MMAP searches for covariates in the following order:

1. Phenotype file
2. Covariate files, if any, in the order listed in the –covariate_filename option.
3. Binary genotype file that contains genotype data.

MMAP uses the first instance of the covariate found. So if BMI is found in the phenotype file, then BMI would be ignored in the covariate files, if any and present. There is no merging of covariate variables across files. Covariates in interaction terms need not be in the same file. SNP covariates in the **--interaction** models can be in covariate file directly or extracted from the binary genotype file using the option **--binary_covariate_filename <file>.** However, there is an important difference between the two options. If the SNP is in the covariate file, then missing genotypes have to be coded as blank or NA, which results in those subjects being excluded from the analysis. If the SNP is extracted from the binary genotype file, then missing data is imputed to the population average of the subjects in the analysis (not genotype file). Thus, the number of individuals analyzed may be different between the two approaches when genotype data is missing. We may add an option to use only observed data in the future, but the option –subject_set with the ids of the individuals with genotype data can be used to restrict the analysis. Also the imputed value for missing data is based on the subjects in the analysis, so the analysis is not impacted if genotype file has mixed ethnicities.

Robust Standard Errors

Linear Regression

For linear regression MMAP implements a menu of heteroskedasticity consistent (HC) estimators HC0 (Huber-White), HC1, HC2, HC3, HC4, HC4m and HC5 as defined in the R **sandwich** package. These estimators model the variance of the beta estimate as $Var(\hat{\beta}) = \left(X'X\right)^{-1} X'\hat{\Omega}X \left(X'X\right)^{-1}$ with diagonal matrix $\hat{\Omega} = diag(w_1, w_2, ..., w_n)$ where the choice of weights w_k is determined by the HC model. The weights are generally a function of the

residuals $e_k = y_k - X_k\hat{\beta}$ and diagonals h_{kk} of the Hat matrix $H = X\left(X'X\right)^{-1}X'$. For example the Huber-White weights are $w_k = \hat{e}_k^2$ and the weights for HC3 are $w_k = \hat{e}_k^2 \big/ \left(1-h_{kk}\right)^2$.

Any combination of the six sandwich estimators can be included in the regression analysis.
--hc0_sandwich_estimator
--hc1_sandwich_estimator
--hc2_sandwich_estimator
--hc3_sandwich_estimator estimator
--hc4_sandwich_estimator
--hc4m_sandwich_estimator
--hc5_sandwich_estimator

The MMAP output will contain HC0, HC1, HC2, HC3, HC4, HC4m and HC5 prefixed columns containing the estimates for the standard errors and the p-values. Note that the betas are not changed by the sandwich estimators. If sandwich estimators are included in the interaction model, the robust covariance estimates are also provided in the <trait>.<prefix>.mle.pval.csv output file.

Mixed model

In the mixed model the variance structure *V* is not the identity, thus $Var(\hat{\beta}) = \left(X'V^{-1}X\right)^{-1}X'V^{-1}\hat{\Omega}V^{-1}X\left(X'X\right)^{-1}$**.** The sandwich $\hat{\Omega}$ that is currently supported is $\hat{\Omega} = \mathrm{diag}(\hat{\varepsilon}\hat{\varepsilon}')$, where $\hat{\varepsilon} = y - X\hat{\beta} = \hat{g} + \hat{e}$ which corresponds to the formula for the Huber-White HC0 estimator. This matrix is diagonal so it models the variance of *Y* treating individuals as independent clusters. There are no hat options at the moment as computing the Hat matrix requires greater computational burden.

This is a work in progress so other formulations that may be more appropriate are being researched, for example, $\hat{\Omega} = \hat{g}\hat{g}' + \mathrm{diag}(\hat{\varepsilon}\hat{\varepsilon}')$. If there is any concern regarding the HC0 model, then an alternative strategy is use MMAP to fit the baseline variance component model, say polygenic model, then use the error residuals in output file as the phenotype in the linear regression (--linear_regression). These residuals are then independent.

The required option is **--empirical_sandwich** and the output file will have HC0 as the header.

Notes:

1. For GxE interaction meta-analysis plans, the robust estimators refer to Huber-White (--HC0 and --empirical_estimator). The properties of other HC estimators for genetic analysis has not been investigated.
2. Currently the calculation of robust estimators is not optimized with MKL libraries.

Example MMAP commands

Linear Model

mmap [ped, trait, covariate, genotype options] **--linear_regression --hc_sandwich_estimator --hc1_sandwich_esimator --hc2_sandwich_sandwich**

Mixed Model

mmap [ped, trait, covariate, genotype options] --read_binary_relationship_matrix <file> **--empirical_sandwich**

Sandwich package: http://cran.r-project.org/web/packages/sandwich/index.html