README - JLS

May 19, 2015

Note that the *current* implementation of the Joint Location-Scale method is appropriate for testing association of a phenotype with a *genotyped* genetic variant (e.g. SNP) or an imputed variant with a 'hard call' (i.e. assign individual to the genotype with the highest posterior imputation probability; high genotype uncertainty does not affect type 1 error but will decrease power), using a sample of *unrelated* subjects. Extension of the method to incorporate genotype uncertainty and deal with correlated subjects and a more user-friendly analytic toolset is underway.

Steps for applying the JLS testing procedure for single-SNP analysis

1. Check the phenotype of interest for fit to a normal distribution. If required, adjust the phenotype using a suitable transformation, e.g. inverse normal transform. If the researcher proceeds using a non-normal phenotype, only the permutation (resampling-based) p-value analysis will be valid (see step 4 (b)).
2. Choose the individual location and scale tests based on the distribution of phenotype (normal or non-normal) or preference, for example parametric or non-parametric versions of each test. Here we choose linear regression and Levene's test for the location and scale tests, respectively.
3. Choose a JLS testing method of combining information from the individual location and scale tests and calculate the JLS test statistic. We note that there is no 'most powerful' method for all situations in practice. Based on our experience, we recommend the use of Fisher's method (JLS-Fisher) of combining the association evidence:

$$W_F = -2\,(\log(p_L) + \log(p_S)),$$

where p_L and p_S are the individual location and scale test p-values, respectively.
4. Chose the p-value estimation method for the JLS statistic.

(a) Based on the approximate asymptotic distribution of the JLS test statistic: For the JLS-Fisher example, W_F is distributed as a χ^2_4 random variable, if the chosen individual location and scale tests are independent of each other under the null hypothesis. This assumption is correct if the trait is normally distributed and if the location-only test statistic is a function of the complete sufficient statistic (e.g. linear regression t-statistic, ANOVA F-statistic) and the distribution of the scale-only test statistic does not depend on the model parameters (e.g. Levene's test or the F-test for equality of variances).

(b) Based on resampling methods such as permutation:

- Calculate the observed JLS test statistic, e.g. W_F.
- Choose the number of permutation replicates, *K*, based on the desired p-value accuracy. In the genome-wide setting, *K* must be large (e.g. 10^7 for p-value in the range of 10^{-5}) and this is computationally expensive.
- Permute the phenotype independently *K* times (not valid if subjects are correlated with each other), and for each replicate *k*, recalculate the JLS test statistic, W_F^k, $k = 1, \ldots, K$.
- Obtain the permutation p-value as {the number of $W_F^K > W_F$}/K.

Steps for applying the JLS testing procedure for gene-set/pathway analysis

5. Annotate *J* SNPs to the gene, gene-set or pathway of interest.
6. Chose the specific single-SNP JLS test (e.g. JLS-Fisher) for each SNP *j*, and obtain the corresponding test statistic, e.g. $W_{F,j}$.
7. Aggregate the association evidence across the *J* SNPs and obtain the sum, $\sum_j W_{F,j}$.
8. To account for LD between SNPs, assess the overall association evidence using a phenotype-permutation approach similar to 4(b).
    - Choose the number of permutation replicates, *K*, based on the desired p-value accuracy. (Because this multivariate method analyzes all *J* SNPs simultaneously, the number of permutations need not be exceedingly large and $K = 10^4$ provides accurate estimates for p-values in the range of 0.05. However, if multiple genes or gene-sets were of interest, more

replicates would be required to adjust for the corresponding number of hypothesis tests.

- Permute the phenotype independently *K* times (not valid if subjects are correlated with each other), and for each replicate *k*, recalculate the JLS *sum* test statistic, $\sum_j \quad W_{F,j}^k$, $k = 1, \ldots, K$.
- Obtain the permutation p-value as {the number of $\sum_j \quad W_{F,j}^k > \sum_j W_{F,j}$}/K.