Accelerated Bivariate Heritability Inference Using the Endophenotype Ranking Value

Xu Chen[1], Neda Jahanshad[2], Paul Thompson[3], David Van Essen[4], Peter Kochunov[5], Thomas Nichols[6]

[1] Department of Cognitive Neuroscience, Maastricht University, Maastricht, Netherlands,
[2] Imaging Genetics Center, Institute for Neuroimaging & Informatics, University of Southern California, Los Angeles, CA, USA,
[3] Keck School of Medicine of USC, Los Angeles, CA, USA, [4] Washington University, St Louis, MO, USA,
[5] Maryland Psychiatric Research Center, Baltimore, MD, USA, [6] Department of Statistics, University of Warwick, Coventry, UK
email: xu.chen@maastrichtuniversity.nl, t.e.nichols@warwick.ac.uk

Introduction

Heritability analyses measure the extent to which the inter-subject variation can be accounted for by the genetic influences. This analysis can be done for a single observable trait (i.e., trait heritability) and also for multiple traits, where heritability concerns the overlapping genetic effects shared between measurable characteristics (i.e., genetic correlation between traits). In OHBM 2013, we proposed a novel linear regression method (LR-SD [1]), using the squared differences of paired observations to infer trait heritability. Here, we generalise the univariate LR-SD method to the bivariate genetic modelling of a pair of phenotypes. The integration of univariate and bivariate LR-SD methods provides a new fast estimation approach for genetic correlation. However, while our initial attempts to estimate genetic correlation were unsuccessful (severe bias), we found that we could accurately estimate a closely related measure, the endophenotype ranking value (ERV). ERV depends on the genetic correlation and the heritability of each phenotype, and can be used to test for zero heritability or genetic correlation. Compared with the commonly used and computationally-intensive likelihood ratio test statistic, the ERV is so fast that permutation inference is practical. We perform simulations to evaluate the validity of this ERV statistic and illustrate it on a real dataset.

Methods

The ERV is defined as ERV $= h_1 h_2 \rho_g$, where h_1 and h_2 are square root of the heritability of two phenotypes, and ρ_g denotes the genetic correlation between phenotypes. Bivariate LR-SD uses only differences squared (DS's) among twin pairs and their two phenotypes, and estimates the ERV using ordinary least squares. The subject pairs can be partitioned into 3 groups: MZ, DZ and UN (unrelated subjects).

The expected value of DS depends on ERV in a simple way, allowing direct estimation; for example, for standardised phenotypes:

DS	$\mathbb{E}$[DS]
Inter-MZtwin, Inter-Pheno	$2 - 2$ERV
Inter-DZtwin, Inter-Pheno	$2 -$ ERV

For hypothesis testing, we use absolute value of the ERV, i.e., $|h_1 h_2 \rho_g|$, to equally consider the positive and negative ERV. The null hypothesis H_0: ERV $= 0$ is equivalent to H_0: $\rho_g = 0$, since zero heritability implies $\rho_g = 0$. Although the null distribution of the ERV is unknown, the permutation test provides a simple way to estimate its exact (empirical) null distribution. When the null hypothesis is true, MZ and DZ twin pairs are exchangeable and can be randomly shuffled.

Twin data from the Human Connectome Project (HCP; [2]) was extracted to form a sample of 178 healthy adults (126 females, 52 males) including 98 MZ and 80 DZ twins from 89 families with an age range of 22-36 (mean$\pm$SD: 29.8$\pm$2.9). Using the approach of the ENIGMA project [3], the raw DTI image of each subject was pre-processed and quality-controlled to derive images of fractional anisotropy (FA). Once the FA images were obtained, the ENIGMA-DTI template was used for image alignment; 36 ROI's were extracted and grouped (based on major white matter tracts), and mean FA for each ROI was calculated as the regional phenotypic measure [3]. The ERV statistic was utilised within the permutation framework to test the significance of combined genetic influences between ROI's. 1000 permutations were used to find ERV p-values.

Results: Simulations

Fig 1 shows the rejection rate (in percent) at level 0.05 for H_0: ERV $= 0$ when $h^2 = 0$ (blue), $h^2 = 0.25$ (green) and $h^2 = 0.5$ (yellow), $n = 50 + 50$ (50 MZ and 50 DZ twins, panel A), $n = 100 + 100$ (panel B) and $n = 200 + 200$ (panel C). 1000 simulations, 1000 permutations each, were performed. The red dash-dotted lines show the lower and upper bounds of the 95% binomial proportion confidence interval. The null hypothesis is true when $h^2 = 0$ or $\rho_g = 0$, which corresponds to all blue bars and the 3 green and yellow bars marked $(0, *)$. For nearly all null simulation settings, the estimated FPR (the rate of falsely rejecting the null hypothesis when it is actually true) lies below the upper bound of the 95% binomial proportion confidence interval, which implies the ERV statistic is valid for assessing the null hypothesis.

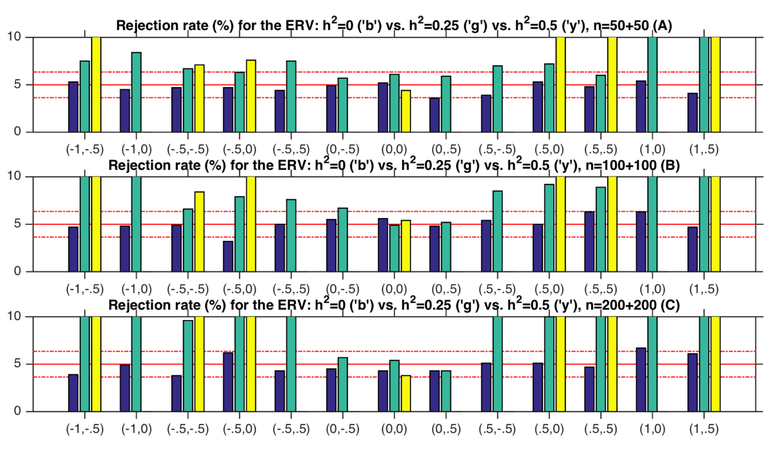

Figure 1: Simulation results: the rejection rate (in percent) at level 0.05 using the ERV for different heritability levels. The x-markers correspond to (ρ_g, ρ_p).

Results: Real Data

Figs 2 & 3 show the significant bivariate genetic correlations for ROI pairs, assessed with the ERV statistic using 1000 permutations after FWE (Fig 2) and FDR (Fig 3) corrections. Each marker on x and y axes corresponds to a ROI. Univariate LR-SD analysis found non-zero heritability for all ROI's, ranging from 0.04 to 0.85. With the use of bivariate LR-SD, the most significant FWE-corrected p-value for those connected ROI pairs is 0.001 with 3 significant ROI pairs, and the best attainable FDR-corrected p-value is 0.014 with 199 significant ROI pairs found. This implicitly demonstrates that the FDR control of p-values for multiple comparisons is more powerful than the FWE correction.

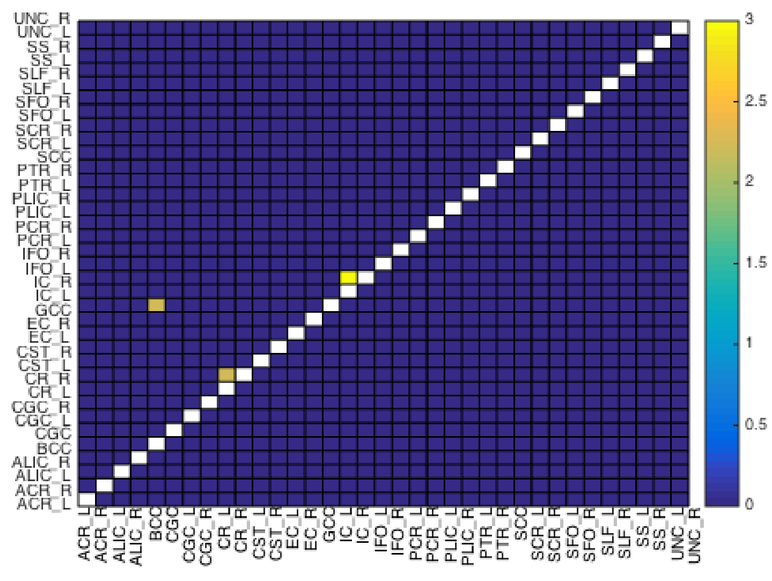

Figure 2: $-\log_{10}$(FWE corrected p-values for ERV).

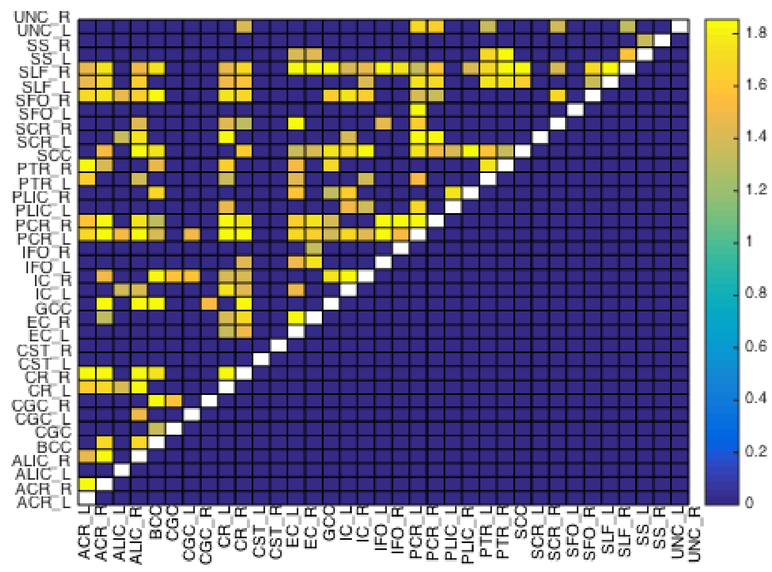

Figure 3: $-\log_{10}$(FDR corrected p-values for ERV).

Conclusions

We have extended the univariate LR-SD method to the bivariate case, suggested a fast test statistic of the ERV for testing zero genetic correlation using this bivariate LR-SD method, and demonstrated the ERV inference with a real data application. With the simulation evaluations, we have shown that the ERV statistic is valid and has reasonable power. Crucially, our fast approach to ERV inference allows the use of the non-parametric permutation inference.

References

[1] Chen, X, et al. (2013), *OHBM*, Poster 1289.
[2] Van Essen, DC, et al. (2013), *Neuroimage*, 80:62-79.
[3] Jahanshad, N, et al. (2013), *Neuroimage*, 81:455-469.