

Estimating kinship matrices in the presence of linkage disequilibrium

*Susana Eyheramendy*¹, *Eleazar Eskin*²

¹susana@mat.puc.cl, Department of Statistics, Pontificia Universidad Catolica de Chile.

²eeskin@cs.ucla.edu, Department of Computer Science, University of California, LA.

Over the past several years, mixed-model methods have been applied widely to the analysis of human genetic data. These methods have been utilised to correct for population structure, to estimate heritability, to estimate genetic correlation in multivariate analysis and trait predictions.

A key step in any mixed-model analysis is estimating genetic relationships from genome-wide genetic variation information. A question that has been debated in the literature is what is the effect of correlation between genetic variants, or linkage disequilibrium, on estimating kinship matrices and how should kinship estimation methods take into account linkage disequilibrium. Many difficulties in estimating the relatedness from SNP data have arisen. These difficulties include issues related to the unknown genetic architecture of a trait such as the unknown relationship between the frequency of variants and their effect sizes, the unknown number of variants with an effect on the trait and their distribution throughout the genome and the unknown distribution of effect sizes among the variants that have an effect. Another class of difficulties are related to the fact that the collected genetic data is incomplete, which includes that the causal variants are unlikely directly present in the collected data. Finally, some have also considered the issue that the variants are correlated or in LD and the effects of this correlation on the estimation of kinship matrices.

Exactly how to estimate the kinship is not a resolved question. In this paper, we revisit the question of how linkage disequilibrium affects kinship estimation. We focus on this issue, and for simplicity, we assume that all variants are collected and the genetic architecture is known. We consider three estimators of kinship matrices where one ignores linkage disequilibrium and the other two incorporates linkage disequilibrium information. The first estimator is the standard genetic relationship estimator of Yang *et al.* The second estimator takes into account linkage disequilibrium, was proposed by Speed *et al.* The third estimator is a new estimator that we propose which also takes into account linkage disequilibrium between SNPs. We provide analytical derivations of the mean and variance of the three estimators and show that they are all unbiased but taking into account linkage disequilibrium results in a more accurate estimate of the kinship matrix.

Keywords: Linear Mixed Models, heritability, kinship matrices