

The effect of population kinship estimation bias in heritability estimation and genetic association

Zhuoran Hou¹, Alejandro Ochoa^{1,2,*}

¹ Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27705, USA

² Duke Center for Statistical Genetics and Genomics, Duke University, Durham, NC 27705, USA

* Corresponding author: alejandro.ochoa@duke.edu

Abstract

Population kinship matrices are estimated for a variety of applications, including estimation of heritability and to control for population structure in genetic association studies. Recent work found that the most common family of kinship estimators can be severely biased. In this work, we investigate the effect of this kinship bias on the two downstream applications of heritability estimation and genetic association. We present a novel trait simulation strategy that accurately parametrizes heritability, even when utilizing real genotypes. Using these simulations, we find that heritability estimation becomes biased when using such biased kinship matrices. Remarkably however, this kinship bias does not affect genetic associations based on either Principal Components Analysis (PCA) or Linear Mixed-effects Models (LMM). Lastly, we explain our empirical observations using theory. In particular, the exact form of the bias of the standard kinship estimator is such that it is compensated for by fitting the intercept in both PCA and LMM approaches, which model population structure via covariates, suggesting that downstream applications without this precise arrangement will not be robust to this kinship bias.

1 Introduction

...

2 Methods

...

3 Results

...

4 Discussion

...