Comparison of Heritability Estimates from Linear Mixed Models and LD-Score Regression

Chhavi Yadav (A59002708), Nick Rittler (A59002648)

March 2024

1 Introduction

Traits are the result of a complex mixture of genetic and environmental effects. The *heritability* h^2 of a trait can be defined as the proportion of phenotyopic variation due to variation in genetic values [2]. Given that e.g. searches for causal genetic factors in phenotypic expression only make sense insofar as a trait is heritable, it is highly desirable that we be able to accurately estimate hertiability.

In this work, we compare two hertiabtilty estimation methods, namely an application of Linear Mixed Models (LMMs) and Linkage Disequilibrium (LD)-Score Regression. In the LMM approach, traits are modeled using single nucleotide polymorphism (SNP)-specific random effects, and estimates of trait heritability $\hat{h}^2$ are given in terms the total estimated variance across SNPs; in LD-Score regression, heritability estimates $\hat{h}^2$ are proportional to the slope of the line of best fit connecting genome wide association study (GWAS) test statistics and LD-scores.

2 Methods

2.1 Description of Data, Phenotype Simulation

We used the 1000 Genomes dataset as the basis for our investigation. We limited our attention to chromosome 21, and individuals from the following populations codes: “CEU” (Northern Europeans from Utah), “PEL” (Peruvian),“YRI” (Yoruba Nigerians), “FIN” (Finnish), “PUR” (Puerto Rican), “CHS” (Southern Han Chinese), “PJL” (Pakistani Punjabi), and “GBR” (Great Britain). This yielded a diverse set of 787 individuals for our study. We further restricted the dataset to SNPs with minor allele frequency (maf) $\geq .1$. This left us with a final dataset containing around 90,000 SNPs for each individual.

The 1000 Genomes dataset does not come with phenotypic data. Thus, to complete our study, we had to simulate artificial traits. Simulations were performed exclusively with the GCTA software, version 1.94.1, following their examples closely [3]

To do this, we specified 10, arbitrarily chosen causal SNPs, and for a given ground truth heritability h^2, simulated the phenotype for each individual y_j according to the following formula $y_j = \sum_i w_{ij} \cdot \gamma_i + e_j,\ e_j \sim N(0, \sigma_j^2)$. Here, $w_{ij} = (x_{ij} - 2p_i)/\sqrt{2p_i(1-p_i)}$, where p_i

represents the maf at the i^{th} causal SNP, and $\sigma_j^2 = var(\sum_i w_{ij} \cdot \gamma_i) \cdot (1/h^2 - 1)$. In all cases γ_i is the effect size of the i^{th} causal SNP.

We considered a variety of heriabilities and effect sizes in our study. We first specified ground truth h^2 values $.01, .1, .2, .5, .9$. We then set a base effect size vector $\gamma = (.01, .05, .1, .2, .5, 1, 1, -1, -1., -.01)^T$, and considered constant multiples of γ for each constant $c \in \{.01, .1, .5, 1, 2, 5, 10\}$. For each pair of effect size vector $c\gamma$ and ground truth h^2, we simulated a trait for each individual, yielding 30 phenotypes for each individual in total.

2.2 Linear Mixed Model (LMM) Approach

The LMM approach to h^2 estimation to assume the phenotypic response vector y (where each entry corresponds to an individual) is generated according to

$$y = C\beta + g + \epsilon,$$

where C is matrix of fixed effects representing controlling factors, $g \sim N(0, A\sigma_g^2)$ for a "genetic relationship matrix" A, and $\epsilon \sim N(0, \sigma_\epsilon^2)$ are "environemntal" errors. The term g corresponds to the random effects for each SNP, with genetic relationship matrix A gives covariance between individuals based on genome similarity [4]. Such models can be fit using restricted maximum likelihood (REML) methods. The estimate for the heritability is estimated proportion of the variance of y that arises from the SNP random effects, i.e $\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_\epsilon^2}$.

In accordance with the above, for each of our 30 experiments, we have a vector of length 787 whose entries represent a trait for each individual in the study. The matrix C was taken to be the 787×3 matrix of projections onto the top 3 principal components (PCs) resulting from a Principal Componenet Analysis (PCA) on the SNP data, performed using plink version v1.90b6.9.

We carried out the fitting of the LMMs with the GCTA software. This is done in two steps in accordance with the tutorial on their website: first, the genetic relationship matrix is constructed, and then REML used to do the parameter estimation [3].

2.3 LD-Score Regression Approach

The essential idea in LD-score regression it to compute χ_i^2 association statistics of each SNP i with the phenotypic response from a GWAS, compute LD-scores, and then regress these statistics on the LD-scores. We assume that these are connected via the following relation: $\chi_i^2 = \frac{Nh^2}{M} l_i + Na + 1 + e_i$, where l_i is the LD-score for SNP i, $N = 787$ the number of individuals in the study and $M = 87568$ is the number of SNPs, and $e_i \sim N(0, \sigma^2)$. The heritability estimate is then $M\beta/N$ for N for β the slope of line of best fit of the regression of test statistics on LD-scores.

We used plink to first carry out the GWAS, GCTA to compute the LD-scores, and then the python package LDSC to complete the LD-score regression given the statistics and LD scores [1].

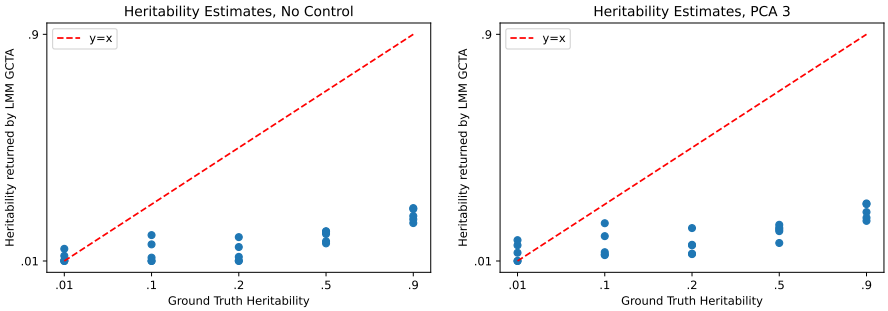

Figure 1: LMM heritability estimates with and without PC controls.

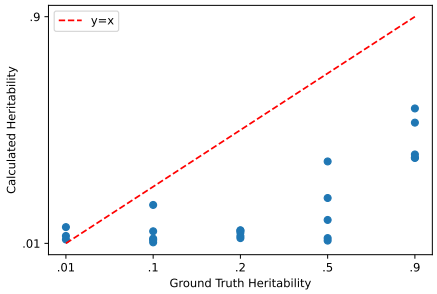

Figure 2: LD-score regression heritability estimates.

3 Results

We found that the LMM approach consistently underestimated heritability, as showcased by the plots in Figure 1. In these plots, the ground truth h^2 can be found on the x-axis. The y-axis encodes the estimated $\hat{h}^2$ returned by GCTA's REML fitting. Each blue dot at a given ground truth h^2 value represents the estimate for a distinct underlying effect size vector $c\gamma$. Ideally, we would want the ground-truth h^2 to approximately match the estimated $\hat{h}^2$, i.e. fall near the $y = x$ line. This however is clearly not the case, with most estimates falling significantly below this line. The estimates were fairly consistent for a given ground-truth h^2 across the different $c\gamma$. Interestingly, controlling for population using PCs did not have a significant effect on the estimates, though it did slightly improve them. We experimented with the number of PCs onto which we project, but did not see significant result changes.

A similar phenomenon was observable with LD-Score Regression, in accordance with Figure 2. LD-Score Regression exhibited a larger variance in heritability estimates, but was in certain cases able to get much closer to the true underlying h^2, particularly at $h^2 = .9$. We did not notice a pattern connecting the underlying effect size and accuracy of $\hat{h}^2$ for LD-score regression; in other words, the lack of consistency in the estimates at a given ground-truth h^2 did not seemed to be related to effect size vectors $c\gamma$. In some cases, heritability estimates from LD-scores were negative, though always small in magnitude in this case. While in principle statistically possible, this is nonsensical, and somewhat of a downside of the method. For plotting purposes, we set $\hat{h}^2$ in these cases to 0.

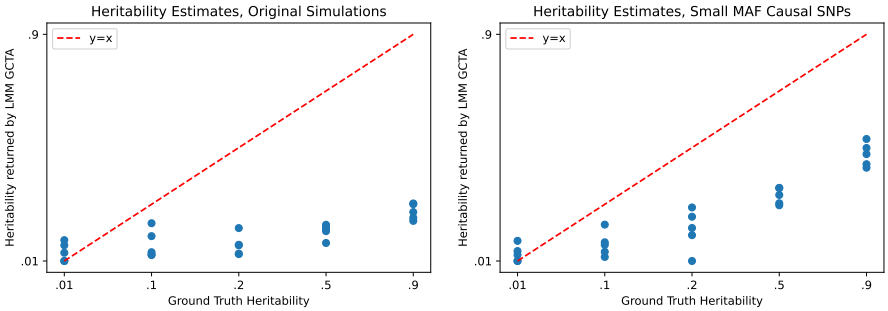

Figure 3: LMM heritability estimates with large and small minor allele frequency causal SNPs. The right panel, corresponding to small minor allele frequency causal variants, showcases improved accuracy of LMM estimates.

4 Discussion

4.1 Potential Reasons for Underestimation Phenomenon

Given LMMs and LD-Score regression are regarded as useful tools for heritability estimation, we were to surprised to see such poor estimation performance, particularly as restriction to a single chromosome with maf $\geq .1$ seemed to only create an easier setting for estimation.

Its unclear to us whether our simulations fail to reflect some important structure which is respected by real-world phenotypic data. In restricting to maf $\geq .1$, and arbitrarily choosing causal SNPs in our simulations, we implicitly created traits whose h^2 was specified by some small amount of SNPs on which many individuals had the minor allele, which is likely an unrealistic setting.

We realized that one culprit might be the specification of the variance for the simulated trait of individual j, namely $\sigma_j^2 = var(\sum_i w_{ij} \cdot \gamma_i) \cdot (1/h^2 - 1)$. Recall that $w_{ij} = (x_{ij} - 2p_i)/\sqrt{2p_i(1-p_i)}$, where p_i represents the maf at the i^{th} causal SNP. Thus, this formula implies that if causal SNPs all have mafs somewhere near $1/2$, the σ_j will be roughly the same *regardless of genotype*. It makes sense that the random effects for each SNP are very hard to separate in this regime, leading to the variance partitioning being skewed to the environmental variance parameter σ_e^2 and underestimation of the h^2.

We tested this conjecture by repeating our experiments, generating a round of new traits using the same ≈ 90k SNPs, but this time choosing new causal SNPs and modifying the effect size vector γ to have it's largest components at the SNPs with maf $\approx .1$. Because of lack of datahub cooperation and the large size of the GWAS, we were not able to estimate the heritability using LD-score regression for this experiment, but we were using LMMs. The results are visualized in Figure 3. The underestimation phenomenon is still present, but is significantly dampened. This suggests that LMMs may significantly underestimate heritability arising from large minor allele frequency SNPs.

4.2 Comparison of Approaches

As we've seen, neither estimation method performed particularly well on our simulated data. Nevertheless, there were some pros and cons of each approach, which we discuss here.

In terms of runtime, GCTA's LMM implementation was significantly faster, taking just a few minutes to fit the 30 models required for our experiments. On the other hand, LD-Score Regression can take up to a day on datahub to complete all 30 experiments. The bottleneck here is the GWAS, which is required to compute the χ^2 statistics.

In general, we note that LMM estimates were more robust – given some ground-truth h^2, the LMM estimates were fairly consistent across effect sizes. This was not true for LD-Score regression, which in the worst case can even return nonsensical, negative $\hat{h}^2$. This could be due to some error propagation in the multiple steps required to obtain the LD-Score regression estimates.

4.3 Future Directions

If we were to work further on this project, we would try to create much more realistic and detailed phenotype simulations. This would hopefully to clear up much of the underestimation effect, though a larger number of individuals in the study would likely help as well.

5 Code Availability

All code used in the final experiments described above can be found via this link to our Github repository.

References

[1] B. Bulik-Sullivan. Ldsc: A command-line tool to estimate heritability and genetic correlations from gwas summary statistics, 2024. Accessed on February 12, 2024.

[2] N. R. Wray. Estimating trait heritability, 2008. Accessed on February 12, 2024.

[3] J. Yang. Gcta software, 2022. Accessed on February 12, 2024.

[4] J. Yang, S. H. Lee, M. Goddard, and P. M. Visscher. Gcta: a tool for genome-wide complex trait analysis. *American journal of human genetics*, 88 1:76–82, 2011.