

Fast and accurate construction of confidence intervals for heritability

November 23, 2015

Regev Schweiger¹, Shachar Kaufman², Reijo Laaksonen^{3,4,5}, Marcus E Kleber⁶, Winfried März^{6,7,8}, Eleazar Eskin^{9,10}, Saharon Rosset², Eran Halperin^{1,11,12}

1. The Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, 6997801, Israel
2. Department of Statistics, Tel Aviv University, Tel Aviv, 6997801, Israel
3. Zora Biosciences, Espoo, Finland
4. Medical School, Tampere University, Tampere, Finland
5. Finnish Clinical Biobank Tampere, University Hospital of Tampere, Finland
6. Vth Department of Medicine (Nephrology, Hypertensiology, Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, University of Heidelberg, Mannheim D-68167, Germany
7. Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz A-8036, Austria
8. Synlab Academy, Synlab Services GmbH, Mannheim D-68165, Germany
9. Department of Computer Science, University of California Los Angeles, Los Angeles, CA 90095, USA
10. Department of Human Genetics, University of California Los Angeles, Los Angeles, CA 90095, USA
11. Molecular Microbiology and Biotechnology Department, Tel Aviv University, Tel Aviv, 6997801, Israel
12. International Computer Science Institute, 1947 Center St., Berkeley, CA 94704, USA

Correspondence should be addressed to Regev Schweiger (schweiger@post.tau.ac.il).

Abstract

Estimation of heritability is fundamental in genetic studies. In recent years, heritability estimation using linear mixed models (LMMs) has gained popularity, because these estimates can be obtained from unrelated individuals collected in genome wide association studies. Typically, heritability estimation under LMMs uses either the maximum likelihood (ML) or the restricted maximum likelihood (REML) approach. Existing methods for the construction of confidence intervals and estimators of standard errors for both ML and REML rely on asymptotic properties. However, these assumptions are often violated due to the bounded parameter space, statistical dependencies, and limited sample size, leading to biased estimates, and inflated or deflated confidence intervals. Here, we show that often the probability that the genetic component is estimated as zero is high even when the true heritability is bounded away from zero, emphasizing the need for accurate confidence intervals. We further show that the estimation of confidence intervals by state-of-the-art methods is highly inaccurate, especially when the true heritability is either relatively low or relatively high. Such biases are present, for example, in estimates of heritability of gene expression in the GTEx study, and of lipid profiles in the LURIC study. We propose a computationally efficient method, *Accurate LMM-Based confidence Intervals* (ALBI), for the estimation of the distribution of the heritability estimator, and for the construction of accurate confidence intervals. Our method can be used as an add-on to existing methods for heritability and variance components estimation, such as GCTA, FaST-LMM, GEMMA, or EMMA. ALBI is available at <http://www.cs.tau.ac.il/~heran/cozygene/software/albi.html>.

Introduction

It has been known for decades that genetic variation accounts for a substantial portion of disease risk. Quantifying this portion, or estimating the heritability, had been traditionally performed using related individuals such as in twin studies [1–3]. Genome-wide association studies (GWAS) have identified thousands of genetic variants that are associated with dozens of common diseases [4, 5]. A natural approach to estimating the heritability from GWAS is to consider the total heritability explained by the identified variants. However, each of these implicated variants explains only a small fraction of the genetic component of the trait, and for many traits, even the sum total of the contributions of all identified genetic variants only explains a fraction of the heritability that was estimated from twin studies [6].

More recently, linear mixed model (LMM) approaches [7–10] have been applied to estimate the heritability from cohorts of unrelated individuals, such as in GWAS [11]. LMMs achieve considerably more accurate estimates because they can utilize all variants, and not just the variants that are statistically significant from a GWAS. Heritability estimates using LMMs are now widely utilized to understand the genetic architecture of complex traits [12]. In these studies, inferences about the genetic architecture of the trait are made by interpreting the heritability estimates.

Because there is statistical uncertainty in the estimation, a typical study uses confidence intervals (CIs) for the heritability rather than point estimates in its analysis. Unfortunately, as we show in this paper, the CIs reported by current LMM approaches are highly inaccurate (see also [13–17]). This is because these approaches make several assumptions about the data, mainly including the assumption that the estimators follow certain asymptotic distributions. In particular, estimators for traits with either significantly high or low heri-

tability are especially biased because the estimates are near the boundaries. Additionally, these CIs may spread beyond the natural boundaries of their parameters, e.g., including negative values for heritability. As a result, these CIs are often inaccurate, difficult to interpret, or lead to erroneous conclusions.

These biases exist because estimators do not necessarily obey the conditions required for them to asymptotically follow the normal distribution. To cope with this, non-standard asymptotic theory for boundary and near-boundary maximum likelihood estimates has been developed for independent data (e.g., [18–20]), while others [21] have extended the theory to the case where the phenotype can be partitioned into sufficiently many independent vectors. Unfortunately, these conditions typically do not hold for genomic datasets. First, only a single observation is typically drawn from the distribution. Second, the parameter space is bounded, as the heritability is only defined between zero and one, and the true heritability value may lie on the boundary or near it. Therefore, an alternative approach is required.

The disagreement between the theoretical distribution of estimators and their empirical distribution results in consequences which are not limited to CIs only. One important property of the theoretical distribution of estimators, when the true parameter value is near the boundary, is a high probability for a boundary estimate [19, 20]. This is in line with a plurality of reports in the literature, in which the heritability of a phenotype is often estimated to be zero or one [14, 17, 22–24] using restricted maximum likelihood (REML) estimators, which are the state-of-the-art method in heritability estimation.

Previous approaches to remedy these issues have taken several directions. Visscher et al. [25] studied the heritability estimator in a range of pedigree- and marker-based experimental designs. While they derived an analytical expression for the variance, their

method assumed that the heritability estimator follows a normal distribution. Inaccuracies in CI coverage probabilities have also been reported in longitudinal studies, leading some authors to suggest exhaustive hierarchical bootstrapping schemes, e.g., [26]. Recently, Furlotte et al. [27] studied the uncertainty in heritability estimates and suggested its quantification using the Bayesian posterior distribution of the heritability value, conditioned on the observed phenotype. Other works have suggested extending the REML estimation method with Bayesian priors, e.g., [22, 24]. It has also been suggested to replace the asymptotic normality assumption with the asymptotics developed for the non-standard boundary case [28]. However, a large sample size is still required for these approximations to be effective. Finally, alternative statistics have been suggested as a basis for building CIs, such as a quadratic function of the estimator for random effects [29], or a ratio of linear combinations of quadratic functions of the phenotype [14, 30]. Unfortunately, all of these methods either assume asymptotics which typically do not hold in practice, or do not pertain to REML estimation.

In this paper, we present a novel approach for accurately building heritability CIs using LMMs. Instead of an asymptotic approximation, our method uses a parametric bootstrap test inversion approach to construct CIs via sampling phenotypes, performing heritability estimation on the sampled phenotypes, and using these estimates as a basis for CI construction [31]. As a naïve implementation of this approach would be computationally prohibitive, we present a highly accurate approximation that allows us to efficiently construct the intervals. We demonstrate our approach on both the Genotype-Tissue Expression [32] and the Ludwigshafen Risk and Cardiovascular Health (LURIC) [33] datasets. An implementation of our approach, which we call Accurate LMM-Based confidence Intervals (ALBI), is available at <http://www.cs.tau.ac.il/~heran/cozygene/software/albi.html>.

Results

Heritability estimates and confidence intervals

Heritability estimates from population samples can be computed from estimates of the parameters of the linear model $\mathbf{y} = \mu + \mathbf{u} + \mathbf{e}$, where μ is the population mean, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{K})$ is the genetic component of the trait, and $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I})$ is the environmental component. \mathbf{K} is a kinship matrix capturing the genetic relatedness between individuals in the sample, constructed from their genotypes. Heritability is defined as $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$. Similarly, under certain assumptions (see Methods), the heritability estimate is $\hat{h}^2 = \hat{\sigma}_g^2 / (\hat{\sigma}_g^2 + \hat{\sigma}_e^2)$. This captures the proportion of the variance of the trait that corresponds to the genetic component.

To quantify the amount of uncertainty in heritability estimates, all current methods report the standard error of their estimators. The main application of these reported standard errors is to imply the construction of CIs for the true genetic and environmental variance components σ_g^2, σ_e^2 , or for the true heritability, h^2 . These CIs are based on the assumption that the estimator is approximately normally distributed. Given an estimated value, \hat{h}^2 , a standard error estimate, \hat{e} , and a required confidence level, $1 - \alpha$ (e.g., 95%), the CI is constructed as $(\hat{h}^2 - k \cdot \hat{e}, \hat{h}^2 + k \cdot \hat{e})$, where $k = \Phi^{-1}(1 - \alpha/2)$, the $(1 - \alpha/2)$ -th quantile of the standard normal distribution. For example, a 95% confidence level CI is computed as $(\hat{h}^2 - 1.96\hat{e}, \hat{h}^2 + 1.96\hat{e})$. The standard error is calculated using the inverse Fisher information matrix of the estimators, or using similar approaches (for details, see e.g., [34]).

Current methods of building confidence intervals are inaccurate

We investigated the accuracy of such CIs, using the standard error estimation performed by GCTA [35] (a popular software for heritability analysis using LMMs), which employs the approach of assuming a normal distribution for the estimator. Suppose we draw a phenotype vector from the the distribution assumed by the LMM. By the definition of a CI with a confidence level of, e.g., 95%, if we were to repeat such a draw multiple times with the same heritability value h^2 , and compute a CI from each draw, then these CIs should cover h^2 95% of the times. For a range of confidence levels (70%, ..., 95%), we used GCTA to build a CI based on the normal approximation for that level. We then checked the percentage of times in which the CI covers h^2 , as a function of h^2 . Figure 1 uses kinship matrices from two real studies, and shows that for a wide range of h^2 values, and particularly for extreme values (small or large), these normal CIs are largely inaccurate. The two studies are the Genotype-Tissue Expression (GTEx) study [32] and the Ludwigshafen Risk and Cardiovascular Health (LURIC) study [33]. The GTEx project is a sample and data resource designed to study the relationship among genetic variation, gene expression, and other molecular phenotypes in multiple human tissues. It provides a collection of multiple different tissues per donor, along with their genotypes. LURIC is targeted to contribute to the identification and assessment of environmental and genetic factors for cardiovascular diseases. Genotypes and lipid profiles are available for patients hospitalized for coronary angiography between 1997 and 2000 at a tertiary care center in Southwestern Germany (see Methods).

The results of Figure 1 imply that the accuracy of the standard normal CIs depends both on the dataset and on the true heritability of the trait. To further demonstrate this, for each value of a true h^2 , we compared the empirical standard deviation of the REML estimator

of \hat{h}^2 to the average of the standard error reported by GCTA. Here too, we found that these standard errors are inaccurate, especially at low and high true h^2 values, with an error of a multiplicative factor of up to $\times 1.7$ (see Figure 2). As a point of comparison, we additionally studied the Northern Finland Birth Cohort study (NFBC) dataset [36], where the extent of these biases is much more limited.

To demonstrate that the above issues are due to the normal approximation for the distribution of the estimator, we repeatedly sampled phenotypes for the studied datasets, assuming a certain heritability, and examined the distribution of the estimators. We observe that, indeed, the distribution is often far from normal, showing skewness and high probabilities for 0 and 1. Figure 3 illustrates this for several values of true h^2 , when $\sigma_p^2 = 1, \beta = \mathbf{0}_p$. In line with the findings presented previously, the distributions from GTEx and LURIC violate the normality assumptions, while NFBC follows the normal distribution more closely.

The high probability of boundary estimates ($\hat{h}^2 = 0$ or $\hat{h}^2 = 1$), evidenced in Figure 3, is strongly reflected in published results from applied research. For example, 55% of the gene expression profiles in the GTEx data were estimated to have heritability of 0, and 15% of profiles to have heritability 1. Similarly, we calculated the heritability of expression profiles of lipids in the LURIC dataset, estimating 42% of the profiles to have $\hat{h}^2 = 0$ (none had $\hat{h}^2 = 1$). Such results are often difficult to interpret. It is not clear whether the prevalence of 0 heritability estimates suggests a true underlying zero value of heritability ($h^2 = 0$), and therefore that a genetic variance component does not exist, or that this estimation is due to statistical noise or possibly a result of an underlying numerical error.

We can use the sampling approach described above to characterize, for a given kinship matrix, the probability that the heritability would be estimated as either 0 or 1 given the true heritability. This probability, as a function of the true heritability, is shown in

Figure 4. As we can see, for the GTEx and LURIC datasets, these probabilities are high for a wide range of heritability values, which explains the large number of observed heritability estimates that equal zero. On the other hand, the NFBC dataset has a low probability of obtaining a zero estimate, unless the true heritability is small (up to ~ 0.15).

Accurate LMM-based confidence intervals

Motivated by these inaccuracies, we have developed a novel method, which we call *Accurate LMM-Based confidence Intervals* (ALBI), for rapidly estimating the distribution of the REML heritability estimator and for computing accurate heritability confidence intervals. We give an overview of the method and its properties here. For the full description see Methods.

An explicit calculation of the estimator distribution, via the parametric bootstrap [37], can be performed as follows. For a given heritability value h^2 , a large number of phenotype vectors is drawn according to the LMM, and the REML estimator \hat{h}^2 is computed for each such vector. This process can be repeated for a grid of feasible h^2 values, as needed, and we can use as many bootstrap replicates as needed in order to achieve a desired accuracy. Unfortunately, this brute-force method is often a prohibitively time-consuming procedure (see Table 1).

To address this, we have developed a fast analytical approximation, which consists of the following elements. First, it depends only on the eigenvalues of the kinship matrix. Second, under several common scenarios, it gives a closed-form formula for the derivative of the restricted log-likelihood function, which can then be easily calculated. This formula also allows for the evaluation of the probability of the estimate falling within an interval without the need to estimate the entire distribution. Finally, all operations are performed using

the eigenvectors of the kinship matrix, which allows for a computational complexity linear in the number of samples. Our approximation is highly efficient (Table 1) and it provides an excellent match to the brute-force estimation (see Methods).

Once we have the estimator distribution for any possible value of the true heritability, we can efficiently construct accurate CIs, based on the duality between hypothesis tests and CIs. For each true value of h^2 , we select a subset of \hat{h}^2 that has sampling probability $1 - \alpha$, according to the respective estimator distribution, and define it to be the acceptance region for that value of h^2 . The CI for a value \hat{h}^2 is the interval containing all values of h^2 for which \hat{h}^2 does not imply the rejection of the hypothesis of the true heritability value being h^2 , with a confidence level of $1 - \alpha$. As a comparison, we show the result of the CIs derived with ALBI (Figure 1). These CIs achieve the desired confidence level accurately.

Using ALBI, and comparing our accurate CIs to those of methods employing the normal approximation, we observed that CIs derived from the normal approximation are often too wide. This is problematic in scenarios where, for example, one is interested to know if the value $h^2 = 0$ is included in the CI, $C_{\hat{h}^2}$. It may then be the case that $\Pr(0 \in C_{\hat{h}^2})$ is estimated to be too high, causing the researcher to infer that a phenotype may have zero heritability, often erroneously. Indeed, $\Pr(0 \in C_{\hat{h}^2})$ is significantly higher in GCTA than in the ALBI CIs, up to $\times 4.1$ times in GTEx (see Figure 5).

A fast preliminary assessment of the accuracy of the normal approximation

As we have demonstrated, the accuracy of the normal approximation of the estimator greatly depends on the kinship matrix of the cohort used to make the estimates. We can use ALBI to efficiently compute the probability of estimating a heritability of 0 or 1, given

the true heritability value, and generate the data shown in Figure 4. ALBI can compute these probabilities without the need to repeatedly sample the phenotypes and without re-estimating the heritability each time, and thus it is highly efficient. Therefore, we can use these estimated probabilities to quantify how accurate we expect the normal approximation of the heritability estimates to be for a given dataset, prior to explicitly estimating the CIs for each of the estimates. For example, looking at Figure 4, it is clear that heritability estimates in the NFBC data will be much more accurate than in the GTEx or LURIC datasets.

Methods

For clarity of presentation, we begin by defining the heritability under the linear mixed model, providing an outline of the REML estimation method, and detailing some important properties of the likelihood function. Finally, we present ALBI, our method of calculating the heritability estimator’s distribution and constructing accurate confidence intervals for heritability.

The linear mixed model

We consider the following standard linear mixed model (see [38] for a detailed review):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where n is the number of samples, \mathbf{y} is a $n \times 1$ vector of random variables, \mathbf{X} is a $n \times p$ matrix of covariates (possibly including an intercept vector $\mathbf{1}_n$ as a first column), $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, \mathbf{Z} is a $n \times m$ design matrix, \mathbf{u} is a $m \times 1$ vector of random

effects, and \mathbf{e} is a vector of errors. We assume \mathbf{u} and \mathbf{e} are statistically independent and distributed normally as $\mathbf{u} \sim \mathcal{N}(\mathbf{0}_m, \frac{1}{m}\sigma_g^2\mathbf{I}_m)$, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}_n, \sigma_e^2\mathbf{I}_n)$. Define $\mathbf{K} = \frac{1}{m}\mathbf{Z}\mathbf{Z}^T$. Under these conditions it follows [11] that:

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma_g^2\mathbf{K} + \sigma_e^2\mathbf{I}_n). \quad (1)$$

Typically, \mathbf{y} is a vector of phenotype measurements for each individual, \mathbf{X} is a matrix of covariates (e.g., an intercept term, sex, age) and \mathbf{Z} is the standardized genotype matrix, i.e., columns have zero mean and unit variance. \mathbf{K} is commonly called the kinship matrix, or the genetic relationship matrix, and is estimated from the genotypes as $\mathbf{K} = \frac{1}{m}\mathbf{Z}\mathbf{Z}^T$ (see A.1).

The heritability is defined as the proportion of total variance explained by genetic factors [39]:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}.$$

Defining $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$, Equation (1) becomes:

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma_p^2(h^2\mathbf{K} + (1 - h^2)\mathbf{I}_n)). \quad (2)$$

Heritability estimation with ML and REML

The most common way of estimating h^2 is REML estimation (see A.2). For completeness, we give an overview of the REML estimation method. REML consists of maximizing the likelihood function that is associated with $n - p$ linearly independent contrasts [40]. The

restricted log-likelihood function is, up to additive and multiplicative constants:

$$\begin{aligned} \ell_{REML}(h^2, \sigma_p^2, \boldsymbol{\beta}) \propto \\ - (n - p) \log \sigma_p^2 - \log |\mathbf{V}_{h^2}| - \log |\mathbf{X}^T \mathbf{V}_{h^2}^{-1} \mathbf{X}| - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}_{h^2}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{\sigma_p^2}, \end{aligned}$$

where $\mathbf{V}_{h^2} = h^2 \mathbf{K} + (1 - h^2) \mathbf{I}_n$. For a fixed h^2 , the values of σ_p^2 and $\boldsymbol{\beta}$ that maximize ℓ_{REML} can be derived analytically, and plugged back in to derive a profile restricted likelihood function f_{REML} , which is a function of h^2 , that is maximized instead.

In Appendix A.3, we show that the distribution of \hat{h}^2 depends only on h^2 , and is invariant under changes to σ_p^2 and $\boldsymbol{\beta}$. We may therefore limit our study to the \hat{h}^2 estimator alone, in the special case of fixed $\sigma_p^2 = 1, \boldsymbol{\beta} = \mathbf{0}_p$, which substantially simplifies the problem.

Estimating the distribution of REML estimators with parametric bootstrap

We now turn to describe a method to estimate the distribution of REML estimators. We begin with a direct calculation, and then describe a faster approximation.

Direct parametric bootstrap

For a fixed value of h^2 (and assuming $\sigma_p^2 = 1, \boldsymbol{\beta} = \mathbf{0}_p$), we can estimate the distribution of \hat{h}^2 with a parametric bootstrap method. Since the distribution of \hat{h}^2 is continuous in the open interval $(0, 1)$, we divide this range into equally sized bins of a fixed size (0.01), and instead estimate the probability mass function of a random variable taking values in the set $\{0, (0, 0.01], (0.01, 0.02], \dots, (0.09, 1), 1\}$.

Explicitly stated, the method consists of the following steps:

1. Random sampling: Draw N (e.g., 10,000) phenotypes $\mathbf{y}_1^*, \dots, \mathbf{y}_N^*$ from the multivariate normal distribution $\mathcal{N}(\mathbf{0}_n, h^2\mathbf{K} + (1 - h^2)\mathbf{I}_n)$.
2. REML estimation: Calculate the REML estimates $\hat{h}^2(\mathbf{y}_1^*), \dots, \hat{h}^2(\mathbf{y}_N^*)$ for each of these phenotypes, using a software package such as GCTA.
3. Density estimation: For each one of the bins above, count the proportion of estimates $\hat{h}^2(\mathbf{y}_i^*)$ that fall in that bin; similarly, compute the fraction of estimates evaluating to a boundary estimate $\hat{h}^2(\mathbf{y}_i^*) = 0$ or 1. Use these fractions as an estimate of the density of \hat{h}^2 for this value of h^2 .

In what follows, we discuss how to perform each one of these steps more efficiently.

Step 1: Random sampling

Drawing a vector \mathbf{y} from the distribution $\mathcal{N}(\mathbf{0}_n, \mathbf{V}_{h^2})$ may be done by drawing a vector of standard i.i.d. normal variables $\tilde{\mathbf{y}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, and calculating $\mathbf{y} = \mathbf{V}_{h^2}^{1/2} \tilde{\mathbf{y}}$. Any statement about \mathbf{y} can then be restated in terms of $\tilde{\mathbf{y}}$, or further in terms of a vector $\mathbf{u} = \mathbf{U}^T \tilde{\mathbf{y}} = \mathbf{U}^T \mathbf{V}_{h^2}^{-1/2} \mathbf{y}$, where \mathbf{U} is a matrix whose columns are the eigenvectors of \mathbf{K} . Note that since \mathbf{U} is orthonormal, \mathbf{u} is also distributed $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. Therefore, instead of drawing multiple phenotypes $\mathbf{y}_1^*, \dots, \mathbf{y}_N^*$, we draw $\mathbf{u}_1, \dots, \mathbf{u}_N \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, and rephrase later stages in terms of these \mathbf{u} -s. Using \mathbf{u} instead of $\tilde{\mathbf{y}}$ simplifies further calculations, and additionally does not require expensive matrix multiplications.

Step 2: REML estimation

Local evaluation. Instead of finding the global maximum of the restricted profile log-likelihood function f_{REML} directly, we employ two changes: (i) we search for local maxima instead of the global maximum; (ii) we use the derivative f'_{REML} instead of f_{REML} itself.

The main advantage of this approach is that in order to evaluate if a point estimate \hat{h}^2 is a maximum, we need only check a local condition at this point. For example, to check if 0 is a local maximum, we simply evaluate if $f'_{REML}(0) \leq 0$, so evaluating all other points \hat{h}^2 across the range is not required. In theory, it is possible that multiple local maxima exist, in which case we compare the likelihood itself directly at these points. However, in practice we have observed that this multiplicity happens only rarely, and even an arbitrary decision between local maxima does not noticeably hurt CI accuracy.

As we are estimating the distribution of \hat{h}^2 at a specified resolution, we are in fact more interested in the question of the estimate \hat{h}^2 being inside an interval. Using the derivative, it is simple to check if an interval (c_1, c_2) contains a local maximum; If $f'_{REML}(c_1) > 0$ and $f'_{REML}(c_2) < 0$, then there exists at least one $\hat{h}^2 \in (c_1, c_2)$ which is a local maximum. While this is only a sufficient condition for the existence of a local maximum within the interval, we find that when the interval is sufficiently small (e.g., of width 0.01), it is also a necessary condition. In addition, the probability of having multiple local maxima inside a small interval is negligible. Therefore, evaluations are only necessary at grid points.

Closed-form formula. From the discussion above, it follows that we are interested in the condition of evaluating the derivative of the profile restricted log-likelihood function for a given phenotype \mathbf{y} , when evaluated at a point \hat{h}^2 , restated as a function of $\mathbf{u} = \mathbf{U}^T \mathbf{V}_{h^2}^{-1/2} \mathbf{y}$.

Fortunately, in several common scenarios, it is possible to give a closed-form expression of this derivative, whose computation is linear-time in the number of individuals.

We describe here the case where \mathbf{K} was estimated from a standardized genotype matrix; and where $\mathbf{X} = \mathbf{1}_n$, i.e. there are no covariates but there is an intercept. This is largely the most common use case. The full derivation for this case, as well as its extension to the case in which \mathbf{X} additionally includes principal components of \mathbf{K} , is given in Appendix A.4. This is a generalization of the work in [41, 42].

Let d_1, \dots, d_n be the eigenvalues of the kinship matrix in decreasing order, \mathbf{K} , and recall that \mathbf{U} is a matrix whose columns are \mathbf{K} 's eigenvectors. For a true heritability value h^2 , a possible point estimate value H^2 , and for $i = 1, \dots, n-1$, define

$$\xi_i^{h^2, H^2} = \frac{h^2(d_i - 1) + 1}{H^2(d_i - 1) + 1} \left(\frac{d_i - 1}{H^2(d_i - 1) + 1} - \frac{1}{n-1} \sum_{j=1}^{n-1} \frac{d_j - 1}{H^2(d_j - 1) + 1} \right),$$

and also define $\xi_n^{h^2, H^2} = 0$.

A necessary condition for $\hat{h}^2(\mathbf{y}) = H^2$, when $0 < H^2 < 1$, is having $f'_{REML}(H^2) = 0$, which translates to

$$\sum_{i=1}^n \xi_i^{h^2, H^2} u_i^2 = 0,$$

where $\mathbf{u} = \mathbf{U}^T \mathbf{V}_{h^2}^{-1/2} \mathbf{y}$ (a proof is given in A.4). Similarly, when $\hat{h}^2(\mathbf{y}) = 0$ or 1, the respective conditions are having $f'_{REML}(0) \leq 0$ or $f'_{REML}(1) \geq 0$, which translate to

$$\sum_{i=1}^n \xi_i^{h^2, 0} u_i^2 \leq 0 \quad \text{or} \quad \sum_{i=1}^n \xi_i^{h^2, 1} u_i^2 \geq 0, \quad (3)$$

respectively. This is in line with the result of [41]. In practice, we only wish to bound the maximum within an interval. A sufficient condition for a local maximum to be in an interval (c_1, c_2) is having

$$\sum_{i=1}^n \xi_i^{h^2, c_1} u_i^2 > 0 \quad \text{and} \quad \sum_{i=1}^n \xi_i^{h^2, c_2} u_i^2 < 0. \quad (4)$$

Step 3: Density estimation

To estimate the distribution of \hat{h}^2 , we draw $\mathbf{u}_1, \dots, \mathbf{u}_N \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$. We then estimate the probability of $\hat{h}^2 = 0$ or 1 as the proportion of \mathbf{u} values for which the respective Equations (3) hold. If the estimation for a certain \mathbf{u}_i is neither 0 or 1, it means the derivative at 0 is positive and the derivative at 1 is negative. Therefore, we can perform a binary search to find a smaller interval (c_1, c_2) in which Equation (4) holds. This can be done until a small enough interval is found, at a specified resolution. In practice, we simply evaluate all intervals in a specified grid (e.g., $(0, 0.01), (0.01, 0.02), \dots, (0.99, 1)$).

A useful feature of this approach is that, in order to estimate the probability of a boundary estimate (e.g., $\hat{h}^2 = 0$) or an interval estimate (e.g., $0 < \hat{h}^2 < 0.01$), we need not find the maximum for each \mathbf{u}_i , instead only checking if Equations (3), (4) hold.

Confidence intervals for h^2

We wish to build confidence intervals with a coverage probability of $1 - \alpha$ (e.g., 95%). The distribution of \hat{h}^2 depends solely on h^2 , so we may assume without loss of generality that $\sigma_p^2 = 1$ and $\beta = \mathbf{0}_p$ (see Appendix A.3). Our approach is based on the duality between hypothesis testing and confidence intervals. For a fixed value h^2 , an *acceptance region* A_{h^2} is defined as the subset of values \hat{h}^2 for which a test does not reject the null hypothesis

that the phenotype is drawn from $\mathcal{N}(\mathbf{0}, \mathbf{V}_{h^2})$. The probability of the event A_{h^2} under $\mathcal{N}(\mathbf{0}, \mathbf{V}_{h^2})$ should be $\geq 1 - \alpha$. This region may be indirectly derived from an actual test (e.g., a generalized likelihood ratio test) or constructed explicitly. The corresponding confidence interval, C_{H^2} , for an estimate $\hat{h}^2 = H^2$ comprises of the set of parameter values for which the estimated values are not rejected, namely:

$$C_{H^2} = \{h^2 \mid H^2 \in A_{h^2}\}.$$

Since the distribution of \hat{h}^2 is bounded and generally asymmetric, the choice of A_{h^2} is not unique. It remains to determine A_{h^2} for every h^2 . We give here a general description of the construction; in Appendix A.5, we give a full description of the method, along with proofs.

Let $c_\beta(h^2)$ be the β -th quantile function of \hat{h}^2 , when the true heritability is h^2 . Since the distribution is of a mixed type with discontinuity points, it may be the case that $\Pr_{h^2}(\hat{h}^2 \in [0, c_\beta]) > \beta$. Specifically, the probability of the interval $[c_{\alpha/2}(h^2), c_{1-\alpha/2}(h^2)]$ might be larger than $(1 - \alpha/2) - \alpha/2 = 1 - \alpha$. We therefore divide our construction into distinct cases.

If there is a range of values $h^2 \in [s, t]$ for which $\Pr_{h^2}(\hat{h}^2 \in [c_{\alpha/2}(h^2), c_{1-\alpha/2}(h^2)]) = 1 - \alpha$, we set

$$A_{h^2} = \begin{cases} [0, c_{1-\alpha}(h^2)] & \text{if } h^2 \in [0, s) \\ [c_{\alpha/2}(h^2), c_{1-\alpha/2}(h^2)] & \text{if } h^2 \in [s, t] \\ [c_\alpha(h^2), 1] & \text{if } h^2 \in (t, 1]. \end{cases}$$

If no such range exists, but for every h^2 , either $[0, c_{1-\alpha}(h^2)]$ or $[c_\alpha(h^2), 1]$ have a probability

of $1 - \alpha$, then set

$$A_{h^2} = \begin{cases} [0, c_{1-\alpha}(h^2)] & \text{if } h^2 \in [0, \delta) \\ [c_\alpha(h^2), 1] & \text{if } h^2 \in [\delta, 1], \end{cases}$$

where δ is a point chosen so that the regions will have the required $1 - \alpha$ probability. Finally, if there exists a value of h^2 for which neither $[0, c_{1-\alpha}(h^2)]$ or $[c_\alpha(h^2), 1]$ have a probability of $1 - \alpha$, we employ randomized confidence intervals to achieve the required accuracy (see Appendix A.5).

Benchmarks

We compared the computational costs of performing a parametric bootstrap procedure to estimate the heritability estimator distribution, using ALBI, the aforementioned GCTA [35] and pylmm [43], a fast and lightweight LMM solver for use in GWAS.

We generated the estimator distribution for $h^2 = 0.5$ with these three methods. For pylmm and GCTA, we generated 100 or 1000 random phenotypes, and estimated the heritability for each phenotype using the respective programs' estimation methods. Using the default flags, GCTA did not converge for $\sim 7\%$ of the samples. For ALBI, we estimated the distributions at a precision of 0.01 and 0.001, using 100 or 1000 random samples.

Timing for all methods did not include calculation of eigenvalues and eigenvectors. Running times are reported for the GTEx, LURIC and NFBC datasets. Runs were aborted after 8 hours. All programs were run on a 2.2Ghz, 64-bit Linux workstation with 64GB RAM.

Datasets

The GTEx dataset

The Genotype-Tissue Expression (GTEx) [32] study is a sample and data resource designed to study the relationship among genetic variation, gene expression, and other molecular phenotypes in multiple human tissues. It provides a collection of multiple different tissues per donor, along with their genotypes. We use whole-genome data, collected by the Illumina HumanOmni5M-Quad BeadChip. Prior to quality control (QC), the data consist of 191 sampled individuals and 4,276,680 SNPs. We apply the recommended QC [32], after which 185 individuals and 3,575,877 SNPs remain. For heritability estimation, we use whole-blood gene expression profiles for 30,116 genes.

The LURIC dataset

The LUdwigshafen RIsk and Cardiovascular Health (LURIC) [33] study is a project contributing to the identification and assessment of environmental and genetic factors for cardiovascular diseases. The study consists of patients hospitalized for coronary angiography between 1997 and 2000 at a tertiary care center in Southwestern Germany. Quality control steps undertaken here include removing samples with $< 95\%$ call rate, SNPs with $< 98\%$ call rate, minor allele frequency of $< 1\%$, or Hardy–Weinberg equilibrium test p -value of $< 10^{-4}$. In addition, individuals for which the reported sex did not match the genotype-inferred sex, as well as individuals whose genotypes were manually observed to be outliers in an MDS plot, were removed. Moreover, from each pair of individuals with relatedness of more than 0.1875, only one was reserved. From these, only individuals for which lipid data were collected were used. This process resulted in 867 sampled individuals

and 687,262 SNPs. For heritability estimation, we use 102 lipid profiles.

The NFBC dataset

We analyze 5,236 individuals from the Northern Finland Birth Cohort (NFBC) data set, which consists of genotypes at 331,476 genotyped SNPs and 10 phenotypes [36]. The 10 phenotypes are triglycerides (TG), high-density lipoproteins (HDL), low-density lipoproteins (LDL), glucose (GLU), insulin (INS), body mass index (BMI), C-reactive protein (CRP) as a measure of inflammation, systolic blood pressure (SBP), diastolic blood pressure (DBP), and height.

Discussion

We have presented ALBI, an efficient method for computing the distribution of the REML estimator of heritability and for constructing accurate confidence intervals. We showed that ALBI is significantly faster than standard parametric bootstrap approaches in computing the true estimator distribution, which otherwise require the explicit construction of random phenotypes and full estimation of heritability for each phenotype. In addition, unlike current methods, ALBI allows the computation of the probability of heritability estimates lying inside a specified interval (and particularly at the boundaries), without the need to estimate the entire distribution.

One of the main limitations of the methods currently used for heritability estimation is that the construction of CIs or standard errors is based on approximations that, as we have shown here, often do not hold in practice, resulting in unreliable CIs. In contrast, the CIs built by ALBI are accurate by construction, and ALBI can be used as an add-on to any of

the current methods (e.g., GCTA [35], GEMMA [10]).

In addition, ALBI may be used as a practical approach to investigate the effect of various kinship matrix operations on the usefulness of the heritability estimator. For example, failing to exclude individuals with high relatedness introduces near-zero eigenvalues to the kinship matrix, whose effect on the estimator distribution can be tested. Similarly, it is possible to test if the common practice of adding the first PCs as fixed effects in the mixed model improves the quality of the estimator. Moreover, while we theoretically expect larger sample sizes to produce smaller CIs, the relationship between sample size and the shape of the spectrum of the kinship matrix is intricate, and must be studied and validated for each dataset individually, which may be done with ALBI.

We focused in this work on the estimation of heritability in the bounded interval $[0, 1]$, which is its natural domain. It is well known that the heritability estimator in this range is biased due to the bounded parameter space (see Figure S1). Some software packages for the estimation of heritability (e.g., GCTA) allow performing the optimization of the maximum likelihood or REML in an unbounded region. The rationale there being that even though a negative heritability is not meaningful, unbounded REML estimates are unbiased [38], and confidence intervals can be used to test whether the heritability is greater than a certain value, which is often the question of interest. However, this approach is problematic; a more natural solution for this task is using a method such as ALBI to compute reliable CIs for the bounded estimates. Furthermore, the domain is not well defined when the resulting kinship matrix is not positive definite.

A promising direction for future research is improving the efficiency of ALBI further. Utilizing efficient interval searches and hashing schemes is expected to improve the time complexity. Additionally, we note that at the core of the approximation is the estimation of the cu-

mulative distribution function (cdf) of a generalized chi-square random variable (for boundary probabilities) or the joint distribution of two such variables (for non-boundary probabilities). Various approximations for these cdfs are available in the context of quadratic forms of normal variables [44], but to our knowledge, none apply to the generalized setting we have presented here.

The method proposed here can be extended in several directions. As the distribution of the heritability estimator is generally asymmetric and of mixed type, there are several ways to define the acceptance regions that are used to determine the CIs. Each such choice comes with its advantages and disadvantages, and it is possible that different choices may be more suitable in different settings. Additionally, we focused here on CIs for \hat{h}^2 only. In practice, researchers may be interested in joint confidence regions for additional parameters, such as one or more of the fixed effects.

Acknowledgements

The authors would like to thank Noah Zaitlen, Eli Levy-Karin and Yael Baran. This study was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University to RS and SK. EH and SR are faculty fellows of the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. RS and EH are partially supported by the Israeli Science Foundation grant 1425/13. SK and SR are partially supported by the Israeli Science Foundation grant 1487/12. RS is supported by the Colton Family Foundation. EH, SK, and RS are also partially supported by National Science Foundation grant III-1217615. EH, RS and EE are supported by National Science Foundation grant 1331176 and United States–Israel Binational Science Foundation Grant 2012304. The study was also funded by a European Union 7th Framework Programme grant number 201668 for the AtheroRemo

Project. EE is supported by National Science Foundation grants 1065276, 1302448, and 1320589 and National Institutes of Health grants R01-MH101782 and R01-ES022282. The data used for the analyses described in this manuscript were obtained from dbGaP accession number phs000424.v4.p1 on 4/4/2015.

References

1. Fisher, R. A. The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433 (1918).
2. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Research* **6**, 399–408 (2003).
3. Macgregor, S., Cornes, B. K., Martin, N. G. & Visscher, P. M. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Human Genetics* **120**, 571–580 (2006).
4. Manolio, T. A., Brooks, L. D. & Collins, F. S. A HapMap harvest of insights into the genetics of common disease. *The Journal of Clinical Investigation* **118**, 1590 (2008).
5. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* **42**, D1001–6 (2014).
6. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).
7. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–23 (2008).
8. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* **42**, 348–54 (2010).

9. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nature Methods* **8**, 833–5 (2011).
10. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**, 821–4 (2012).
11. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565–9 (2010).
12. Kostem, E. & Eskin, E. Improving the accuracy and efficiency of partitioning heritability into the contributions of genomic regions. *The American Journal of Human Genetics* **92**, 558–564 (2013).
13. Lohr, S. L. & Divan, M. Comparison of confidence intervals for variance components with unbalanced data. *Journal of Statistical Computation and Simulation* **58**, 83–97 (1997).
14. Burch, B. D. Comparing pivotal and REML-based confidence intervals for heritability. *Journal of Agricultural, Biological, and Environmental Statistics* **12**, 470–484 (2007).
15. Burch, B. D. Assessing the performance of normal-based and REML-based confidence intervals for the intraclass correlation coefficient. *Computational Statistics & Data Analysis* **55**, 1018–1028 (2011).
16. Kraemer, K. *Confidence intervals for variance components and functions of variance components in the random effects model under non-normality*. PhD thesis (Iowa State University, 2012).
17. Kruijer, W. *et al.* Marker-based estimation of heritability in immortal populations. *Genetics* **199**, 379–398 (2015).
18. Chernoff, H. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 573–578 (1954).

19. Moran, P. A. Maximum-likelihood estimation in non-standard conditions. **70**, 441–450 (1971).
20. Self, S. G. & Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610 (1987).
21. Stram, D. O. & Lee, J. W. Variance components testing in the longitudinal mixed effects model. *Biometrics*, 1171–1177 (1994).
22. Wolfinger, R. D. & Kass, R. E. Nonconjugate Bayesian analysis of variance component models. *Biometrics* **56**, 768–774 (2000).
23. Li, H. & Lahiri, P. An adjusted maximum likelihood method for solving small area estimation problems. *Journal of multivariate analysis* **101**, 882–892 (2010).
24. Chung, Y., Rabe-hesketh, S., Gelman, A., Dorie, V. & Liu, J. Avoiding boundary estimates in linear mixed models through weakly informative priors. *Berkeley Preprints*, 1–30 (2011).
25. Visscher, P. M. & Goddard, M. E. A general unified framework to assess the sampling variance of heritability estimates using pedigree or marker-based relationships. *Genetics* **199**, 223–232 (2015).
26. Thai, H. T., Mentré, F., Holford, N. H. G., Veyrat-Follet, C. & Comets, E. A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. *Pharmaceutical Statistics* **12**, 129–140 (2013).
27. Furlotte, N. A., Heckerman, D. & Lippert, C. Quantifying the uncertainty in heritability. *Journal of Human Genetics* **59**, 269–275 (2014).
28. Stern, S. & Welsh, A. Likelihood inference for small variance components. *The Canadian Journal of Statistics* **28**, 517–532 (2000).

29. Harville, D. A. & Fenech, A. P. Confidence intervals for a variance ratio, or for heritability, in an unbalanced mixed linear model. *Biometrics*, 137–152 (1985).
30. Burch, B. D. & Iyer, H. K. Exact confidence intervals for a variance ratio (or heritability) in a mixed linear model. *Biometrics*, 1318–1333 (1997).
31. Carpenter, J. & Bithell, J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* **19**, 1141–1164 (2000).
32. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multi-tissue gene regulation in humans. *Science* **348**, 648–660 (2015).
33. Winkelmann, B. R. *et al.* Rationale and design of the LURIC study—a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics* **2**, S1–73 (2001).
34. Wasserman, L. *All of statistics: a concise course in statistical inference* (Springer Science & Business Media, 2013).
35. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76–82 (2011).
36. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genetics* **41**, 35–46 (2009).
37. Efron, B. & Tibshirani, R. J. *An introduction to the bootstrap* (CRC press, 1994).
38. Searle, S. R., Casella, G. & McCulloch, C. E. *Variance components* (John Wiley & Sons, 2009).
39. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era—concepts and misconceptions. *Nature Reviews Genetics* **9**, 255–266 (2008).
40. Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).

41. Crainiceanu, C., Ruppert, D. & Vogelsang, T. Probability that the MLE of a variance component is zero with applications to likelihood ratio tests. *Unpublished manuscript* (2002).
42. Crainiceanu, C. M. & Ruppert, D. Proofs of theorems for the JRSS-B paper “Likelihood ratio tests in linear mixed models with one variance component”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**, 165–185 (2004).
43. Furlotte, N. A. & Eskin, E. Efficient multiple trait association and estimation of genetic correlation using the matrix-variate linear mixed-model. *Genetics* **200**, 59–68 (2015).
44. Mathai, A. M. & Provost, S. B. Quadratic forms in random variables: theory and applications. *Journal of the American Statistical Association* (1992).

A Appendix

A.1 Estimation of the kinship matrix

We follow the procedure described in the GCTA software package [1]. Let \mathbf{Z} be a standardized genotype matrix with the i, j -th element $z_{i,j} = (g_{i,j} - 2p_j) / \sqrt{2p_j(1-p_j)}$, where $g_{i,j}$ is the number of copies of the reference allele for the j -th SNP in the i -th sample, and p_j is the frequency of the reference allele, estimated from all non-missing SNPs. The i_1, i_2 -th element of the kinship matrix is:

$$\mathbf{K}_{i_1, i_2} = \frac{1}{|S_{i_1, i_2}|} \sum_{j \in S_{i_1, i_2}} z_{i_1, j} z_{i_2, j} = \frac{1}{|S_{i_1, i_2}|} \sum_{j \in S_{i_1, i_2}} \frac{(g_{i_1, j} - 2p_j)(g_{i_2, j} - 2p_j)}{2p_j(1-p_j)}$$

where S_{i_1, i_2} is the set of autosomal SNPs whose values are non-missing for both individuals i_1 and i_2 . When there are no missing values, it follows that $\mathbf{K} = \frac{1}{m} \mathbf{Z} \mathbf{Z}^T$.

A.2 Estimating heritability, ML & REML

For the sake of completeness, we present the standard derivation of ML and REML estimation [2–4].

A.2.1 Maximum likelihood estimation

Define $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$ and $h^2 = \sigma_g^2 / \sigma_p^2$. Up to additive and multiplicative constants, the log-likelihood function is

$$\ell_{ML}(h^2, \sigma_p^2, \boldsymbol{\beta}) \propto -n \log \sigma_p^2 - \log |\mathbf{V}_{h^2}| - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}_{h^2}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{\sigma_p^2},$$

where $\mathbf{V}_{h^2} = h^2\mathbf{K} + (1 - h^2)\mathbf{I}_n$. Therefore, $\text{Cov}[\mathbf{y}] = \sigma_p^2\mathbf{V}_{h^2}$. For a fixed h^2 , the values of σ_p^2 and β that maximize ℓ_{ML} can be derived analytically:

$$\begin{aligned}\hat{\beta}(h^2) &= (\mathbf{X}^T\mathbf{V}_{h^2}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}_{h^2}^{-1}\mathbf{y} \\ \hat{\sigma}_p^2(h^2) &= \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\beta}(h^2))^T\mathbf{V}_{h^2}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}(h^2))\end{aligned}$$

The problem of maximizing $\ell_{ML}(h^2, \sigma_p^2, \beta)$ therefore reduces to maximizing the profile log-likelihood function $\ell_{ML}(h^2, \hat{\sigma}_p^2(h^2), \hat{\beta}(h^2))$ over h^2 , which is, up to constants,

$$f_{ML}(h^2) = -\log |\mathbf{V}_{h^2}| - n \log (\mathbf{y}^T \mathbf{P}_{h^2}^T \mathbf{V}_{h^2}^{-1} \mathbf{P}_{h^2} \mathbf{y}),$$

where $\mathbf{P}_{h^2} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{V}_{h^2}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}_{h^2}^{-1}$.

A.2.2 Restricted maximum likelihood estimation

REML (Restricted Maximum Likelihood) was introduced to take into account the loss in degrees of freedom due to estimation of fixed effects [3, 5]. REML consists of maximizing the likelihood function that is associated with $n - p$ linearly independent contrasts. The restricted log-likelihood function is, up to constants:

$$\ell_{REML}(h^2, \sigma_p^2, \beta) \propto -(n - p) \log \sigma_p^2 - \log |\mathbf{V}_{h^2}| - \log |\mathbf{X}^T\mathbf{V}_{h^2}^{-1}\mathbf{X}| - \frac{(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}_{h^2}^{-1} (\mathbf{y} - \mathbf{X}\beta)}{\sigma_p^2},$$

where now

$$\hat{\sigma}_p^2(h^2) = \frac{1}{n - p}(\mathbf{y} - \mathbf{X}\hat{\beta}(h^2))^T\mathbf{V}_{h^2}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}(h^2)).$$

A similar analysis gives a profile function of:

$$f_{REML}(h^2) = -\log |\mathbf{V}_{h^2}| - \log |\mathbf{X}^T \mathbf{V}_{h^2}^{-1} \mathbf{X}| - (n-p) \log (\mathbf{y}^T \mathbf{P}_{h^2}^T \mathbf{V}_{h^2}^{-1} \mathbf{P}_{h^2} \mathbf{y})$$

A.3 Invariance of heritability estimates

We prove an invariance property of the likelihood and restricted likelihood functions, which enables us to focus solely on the estimator of h^2 and disregard σ_p^2 and β . For a fixed value of h^2 , the estimation problem reduces to standard regression with a covariance matrix that is known up to a scaling factor, σ_p^2 . This leads to the generalized least squares solution for $\hat{\beta}$, as it is independent of scale. It can be easily shown that, for $\lambda \in \mathbb{R}$ and $\gamma \in \mathbb{R}^p$:

$$\ell_*(h^2, \sigma_p^2, \beta; \mathbf{y}) = \ell_*(h^2, \gamma^2 \sigma_p^2, \lambda \beta + \gamma; \lambda \mathbf{y} + \mathbf{X} \gamma) + \frac{C}{2} \log(\lambda^2),$$

where ℓ_* may be either ℓ_{ML} or ℓ_{REML} , and the constant C is n or $n-p$, respectively. Therefore, for $\lambda \in \mathbb{R}$ and $\gamma \in \mathbb{R}^p$:

$$\begin{aligned} \hat{h}^2(\mathbf{y}) &= \hat{h}^2(\lambda \mathbf{y} + \mathbf{X} \gamma) \\ \lambda^2 \cdot \hat{\sigma}_p^2(\mathbf{y}) &= \hat{\sigma}_p^2(\lambda \mathbf{y} + \mathbf{X} \gamma) \\ \lambda \cdot \hat{\beta}(\mathbf{y}) + \gamma &= \hat{\beta}(\lambda \mathbf{y} + \mathbf{X} \gamma). \end{aligned} \tag{5}$$

If \mathbf{y} is drawn from the distribution of Equation (2) with parameters $(h^2, 1, \mathbf{0}_p)$, then $\mathbf{y}' = \sqrt{\sigma_p^2} \mathbf{y} + \mathbf{X} \beta$ follows the distribution with parameters (h^2, σ_p^2, β) . Therefore, from (5), the joint distribution of $(\hat{h}^2, \hat{\sigma}_p^2, \hat{\beta})$, for any fixed (h^2, σ_p^2, β) , is determined exactly by the joint distribution of these estimators for $(h^2, 1, \mathbf{0}_p)$. An important conclusion is that the distribution of \hat{h}^2 depends only on h^2 . We may therefore limit our study to the \hat{h}^2 estimator

alone, in the special case of fixed $\sigma_p^2 = 1$ and $\beta = \mathbf{0}_p$.

A.4 The distribution of \hat{h}^2

In [6], a necessary condition for estimating $\sigma_g^2 = 0$ is derived, as the condition of the likelihood function having a non-increasing derivative at $\sigma_g^2 = 0$. In addition, they approximate the probability of this event. We extend their work to derive a necessary condition for the ML/REML estimator evaluating to a general $\hat{h}^2 = H^2$. We then use this result to estimate the distribution of \hat{h}^2 .

A.4.1 Condition for estimating $\hat{h}^2 = H^2$, when $0 < H^2 < 1$

Let $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ be the eigen-decomposition of \mathbf{K} , where \mathbf{D} is a diagonal matrix with the eigenvalues d_1, \dots, d_n , and \mathbf{U} is an orthonormal matrix with the eigenvectors as its columns (If $d_i \leq 0$ (due to numeric errors), we round $d_i = 10^{-10}$ to make sure \mathbf{K} is positive definite). Let

$$\mathbf{A}^{h^2, H^2} = \mathbf{V}_{h^2}^{1/2} \mathbf{P}_{H^2}^T \mathbf{V}_{H^2}^{-1} (\mathbf{K} - \mathbf{I}) \mathbf{V}_{H^2}^{-1} \mathbf{P}_{H^2} \mathbf{V}_{h^2}^{1/2}$$

and

$$\mathbf{B}^{h^2, H^2} = \mathbf{V}_{h^2}^{1/2} \mathbf{P}_{H^2}^T \mathbf{V}_{H^2}^{-1} \mathbf{P}_{H^2} \mathbf{V}_{h^2}^{1/2},$$

where \mathbf{P}_{H^2} is defined in Section A.2.1. Let w_1, \dots, w_n be the eigenvalues of $\mathbf{K}\mathbf{P}_0$, where $\mathbf{P}_0 = \mathbf{P}_{h^2}|_{h^2=0}$. Let p_i be an indicator variable determining if the i -th eigenvalue of \mathbf{P}_0 is

0. Define

$$\mathbf{C}^{h^2, H^2} = \mathbf{A}^{h^2, H^2} - \frac{1}{n} \sum_{j=1}^n \frac{d_j - 1}{H^2(d_j - 1) + 1} \cdot \mathbf{B}^{h^2, H^2}$$

and

$$\mathbf{D}^{h^2, H^2} = \mathbf{A}^{h^2, H^2} - \frac{1}{n - p} \sum_{j=1}^n \frac{p_j(w_j - 1)}{H^2(w_j - 1) + 1} \cdot \mathbf{B}^{h^2, H^2}.$$

In addition, let $\psi_i^{h^2, H^2}$ and $\xi_i^{h^2, H^2}$ be the eigenvalues of $\mathbf{U}\mathbf{C}^{h^2, H^2}\mathbf{U}^T$ and $\mathbf{U}\mathbf{D}^{h^2, H^2}\mathbf{U}^T$, respectively, for $i = 1, \dots, n$. Finally, let $\mathbf{u} = \mathbf{U}^T(\mathbf{V}_{h^2})^{-1/2}\mathbf{y}$. Then, the following holds:

Proposition 1. For $0 < H^2 < 1$, a necessary condition for $\hat{h}^2 = H^2$ for \mathbf{y} is:

$$\begin{aligned} \sum_{i=1}^n \psi_i^{h^2, H^2} u_i^2 &= 0 & (\text{for } ML) \\ \sum_{i=1}^n \xi_i^{h^2, H^2} u_i^2 &= 0 & (\text{for } REML) \end{aligned}$$

Proof. A necessary condition for $\hat{h}^2 = H^2$ (when $0 < H^2 < 1$) being a global maximum for $f_*(h^2)$ (where f_* is either f_{ML} or f_{REML}) is $\frac{\partial f_*}{\partial h^2}(H^2) = 0$. We show that this is equivalent to the equations above.

Utilizing Equations (7, 17) from [6], and applying the chain rule to convert these equations

to be a function of h^2 instead of σ_g^2/σ_e^2 , we get:

$$\begin{aligned}\frac{\partial}{\partial h^2} \log |\mathbf{V}_{h^2}| &= \sum_{j=1}^n \frac{d_j - 1}{h^2(d_j - 1) + 1} \\ \frac{\partial}{\partial h^2} (\log |\mathbf{V}_{h^2}| + \log |\mathbf{X}^T \mathbf{V}_{h^2}^{-1} \mathbf{X}|) &= \sum_{j=1}^n \frac{p_j(w_j - 1)}{h^2(w_j - 1) + 1}.\end{aligned}$$

It is straightforward to verify that

$$\begin{aligned}\frac{\partial f_{ML}}{\partial h^2} &= n \cdot \frac{\mathbf{y}^T \mathbf{P}_{h^2}^T \mathbf{V}_{h^2}^{-1} (\mathbf{K} - \mathbf{I}) \mathbf{V}_{h^2}^{-1} \mathbf{P}_{h^2} \mathbf{y}}{\mathbf{y}^T \mathbf{P}_{h^2}^T \mathbf{V}_{h^2}^{-1} \mathbf{P}_{h^2} \mathbf{y}} - \sum_{j=1}^n \frac{d_j - 1}{h^2(d_j - 1) + 1} \\ \frac{\partial f_{REML}}{\partial h^2} &= (n - p) \cdot \frac{\mathbf{y}^T \mathbf{P}_{h^2}^T \mathbf{V}_{h^2}^{-1} (\mathbf{K} - \mathbf{I}) \mathbf{V}_{h^2}^{-1} \mathbf{P}_{h^2} \mathbf{y}}{\mathbf{y}^T \mathbf{P}_{h^2}^T \mathbf{V}_{h^2}^{-1} \mathbf{P}_{h^2} \mathbf{y}} - \sum_{j=1}^n \frac{p_j(w_j - 1)}{h^2(w_j - 1) + 1}.\end{aligned}$$

Let $\tilde{\mathbf{y}} = (\mathbf{V}_{h^2})^{-1/2} \mathbf{y}$. Then, by equating the derivatives to 0, it follows that a necessary condition for achieving a local maximum at H^2 is

$$\begin{aligned}\tilde{\mathbf{y}}^T \left(\mathbf{A}^{h^2, H^2} - \frac{1}{n} \sum_{j=1}^n \frac{d_j - 1}{H^2(d_j - 1) + 1} \cdot \mathbf{B}^{h^2, H^2} \right) \tilde{\mathbf{y}} &= \tilde{\mathbf{y}}^T \mathbf{C}^{h^2, H^2} \tilde{\mathbf{y}} = 0 \quad (\text{for ML}) \\ \tilde{\mathbf{y}}^T \left(\mathbf{A}^{h^2, H^2} - \frac{1}{n - p} \sum_{j=1}^n \frac{p_j(w_j - 1)}{H^2(w_j - 1) + 1} \cdot \mathbf{B}^{h^2, H^2} \right) \tilde{\mathbf{y}} &= \tilde{\mathbf{y}}^T \mathbf{D}^{h^2, H^2} \tilde{\mathbf{y}} = 0 \quad (\text{for REML})\end{aligned}$$

The eigenvalues of \mathbf{C}^{h^2, H^2} and \mathbf{D}^{h^2, H^2} are denoted by $\psi_i^{h^2, H^2}$ and $\xi_i^{h^2, H^2}$, respectively ($i = 1, \dots, n$). Let $\mathbf{u} = \mathbf{U}^T \tilde{\mathbf{y}} = \mathbf{U}^T (\mathbf{V}_{h^2})^{-1/2} \mathbf{y}$. Then the conditions above are equivalent

to:

$$\sum_{i=1}^n \psi_i^{h^2, H^2} u_i^2 = 0 \quad (\text{for ML})$$

$$\sum_{i=1}^n \xi_i^{h^2, H^2} u_i^2 = 0 \quad (\text{for REML})$$

□

A.4.2 Condition for estimating $\hat{h}^2 = 0$ or $\hat{h}^2 = 1$

The following proposition shows that at the boundaries, the equality in the condition from Proposition 1 is replaced with an inequality.

Proposition 2. *A necessary condition for $\hat{h}^2 = 0$ is*

$$\sum_{i=1}^n \psi_i^{h^2, 0} u_i^2 \leq 0 \quad (\text{for ML})$$

$$\sum_{i=1}^n \xi_i^{h^2, 0} u_i^2 \leq 0 \quad (\text{for REML})$$

Similarly, a necessary condition for $\hat{h}^2 = 1$ is

$$\sum_{i=1}^n \psi_i^{h^2, 1} u_i^2 \geq 0 \quad (\text{for ML})$$

$$\sum_{i=1}^n \xi_i^{h^2, 1} u_i^2 \geq 0 \quad (\text{for REML})$$

Proof. For the case $\hat{h}^2 = 0$, the requirement $\frac{\partial f_*}{\partial h^2}(H^2) = 0$ is replaced with $\frac{\partial f_*}{\partial h^2}(0) \leq 0$, leading to the conditions above. Similarly, for the case $\hat{h}^2 = 1$, the requirement is $\frac{\partial f_*}{\partial h^2}(1) \geq 0$. □

A.4.3 Condition for \hat{h}^2 falling in a given interval

For $\hat{h}^2 \in (0, 1)$, we will be estimating the distribution of \hat{h}^2 at a specified resolution. Therefore, we wish to know whether the estimate \hat{h}^2 falls inside an interval (c_1, c_2) . Using the derivative, it is simple to give a sufficient condition for an interval to contain a local maximum.

Proposition 3. *A sufficient condition for the existence of a local maximum in (c_1, c_2) is*

$$\sum_{i=1}^n \psi_i^{h^2, c_1} u_i^2 > 0 \quad \text{and} \quad \sum_{i=1}^n \psi_i^{h^2, c_2} u_i^2 < 0$$

for ML estimation, and

$$\sum_{i=1}^n \xi_i^{h^2, c_1} u_i^2 > 0 \quad \text{and} \quad \sum_{i=1}^n \xi_i^{h^2, c_2} u_i^2 < 0$$

for REML estimation.

Proof. If $f'_{REML}(c_1) > 0$ and $f'_{REML}(c_2) < 0$, then there exists at least one $\hat{h}^2 \in (c_1, c_2)$ with $f'_{REML}(\hat{h}^2) = 0$ and $f''_{REML}(c_1) < 0$. Therefore, a local maximum of f_{REML} exists in (c_1, c_2) . The case for ML is identical. \square

A.4.4 Linear-time simple formula for a class of \mathbf{X} -s

The derivations above hold for a general covariate matrix \mathbf{X} . Here, we limit ourselves to \mathbf{X} -s whose columns are eigenvectors of \mathbf{K} . Let $\mathbf{p} = (p_1, \dots, p_n)$ be an indicator vector specifying if the i -th eigenvector of \mathbf{K} is not a column of \mathbf{X} . This is a generalization of three common special cases:

1. $\mathbf{X} = \mathbf{0}$. Namely, there are no fixed effects. The requirement about \mathbf{X} holds trivially.

Here, $\mathbf{p} = \mathbf{1}_n$.

2. **Intercept.** In many cases, $\mathbf{X} = \mathbf{1}_n$, an intercept vector (e.g., the default in GCTA).

Since each column of \mathbf{Z} is standardized, $\mathbf{Z}^T \mathbf{1}_n = \mathbf{0}_m$, so $\mathbf{1}_n$ is an eigenvector of \mathbf{K} , corresponding to an eigenvalue of 0. Since \mathbf{K} is non-negative definite, this is the lowest eigenvalue, i.e. $d_n = 0$. In practice, this hold only approximately due to numerical errors in computing the eigenvalues, but the approximation is good enough for this analysis to be useful. Here, $\mathbf{p} = (\overbrace{1, \dots, 1}^{n-1}, 0)^T$.

3. **First PCs.** In addition to an intercept term, a common practice is adding to \mathbf{X} the

q largest principal components, corresponding to the first q eigenvectors of \mathbf{K} . Here, $\mathbf{p} = (\overbrace{0, \dots, 0}^q, \overbrace{1, \dots, 1}^{n-q-1}, 0)^T$.

In these cases, it is possible to present a simpler formula.

Proposition 4. *If the columns of \mathbf{X} are eigenvectors of \mathbf{K} , then*

$$\begin{aligned}\psi_i^{h^2, H^2} &= p_i \cdot \frac{h^2(d_i - 1) + 1}{H^2(d_i - 1) + 1} \left(\frac{d_i - 1}{H^2(d_i - 1) + 1} - \frac{1}{n} \sum_{j=1}^n \frac{d_j - 1}{H^2(d_j - 1) + 1} \right) \\ \xi_i^{h^2, H^2} &= p_i \cdot \frac{h^2(d_i - 1) + 1}{H^2(d_i - 1) + 1} \left(\frac{p_i(d_i - 1)}{H^2(d_i - 1) + 1} - \frac{1}{n - p} \sum_{j=1}^n \frac{p_j(d_j - 1)}{H^2(d_j - 1) + 1} \right).\end{aligned}$$

Proof. Recall that $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ is the eigen-decomposition of \mathbf{K} , with the eigenvalues d_1, \dots, d_n . It follows that $\mathbf{V}_{h^2} = \mathbf{U}(h^2\mathbf{D} + (1 - h^2)\mathbf{I}_n)\mathbf{U}^T$. \mathbf{P}_{h^2} is the projection matrix to the subspace orthogonal to the subspace spanned by the columns of \mathbf{X} . Therefore, $\mathbf{P}_{h^2} = \mathbf{U} \cdot \text{diag}(\mathbf{p}) \cdot \mathbf{U}^T$ (and thus this definition of p_i coincides with the one above).

Recall also that

$$\begin{aligned}\mathbf{A}^{h^2, H^2} &= \mathbf{V}_{h^2}^{1/2} \mathbf{P}_{H^2}^T \mathbf{V}_{H^2}^{-1} (\mathbf{K} - \mathbf{I}) \mathbf{V}_{H^2}^{-1} \mathbf{P}_{H^2} \mathbf{V}_{h^2}^{1/2} \\ \mathbf{B}^{h^2, H^2} &= \mathbf{V}_{h^2}^{1/2} \mathbf{P}_{H^2}^T \mathbf{V}_{H^2}^{-1} \mathbf{P}_{H^2} \mathbf{V}_{h^2}^{1/2}\end{aligned}$$

and denote their respective eigenvalues a_i and b_i . Since all the matrices in these expressions are diagonalizable by \mathbf{U} ,

$$\begin{aligned}a_i &= \frac{(h^2(d_i - 1) + 1)(d_i - 1)p_i^2}{(H^2(d_i - 1) + 1)^2} \\ b_i &= \frac{(h^2(d_i - 1) + 1)p_i^2}{H^2(d_i - 1) + 1}.\end{aligned}$$

It follows that

$$\begin{aligned}\psi_i^{h^2, H^2} &= a_i - \frac{1}{n} \sum_{j=1}^n \frac{d_j - 1}{H^2(d_j - 1) + 1} \cdot b_i \\ &= p_i \cdot \frac{h^2(d_i - 1) + 1}{H^2(d_i - 1) + 1} \left(\frac{d_i - 1}{H^2(d_i - 1) + 1} - \frac{1}{n} \sum_{j=1}^n \frac{d_j - 1}{H^2(d_j - 1) + 1} \right) \\ \xi_i^{h^2, H^2} &= a_i - \frac{1}{n - p} \sum_{j=1}^n \frac{p_j(w_j - 1)}{h^2(w_j - 1) + 1} \cdot b_i \\ &= p_i \cdot \frac{h^2(d_i - 1) + 1}{H^2(d_i - 1) + 1} \left(\frac{p_i(d_i - 1)}{H^2(d_i - 1) + 1} - \frac{1}{n - p} \sum_{j=1}^n \frac{p_j(d_j - 1)}{H^2(d_j - 1) + 1} \right),\end{aligned}$$

where we used $p_i^2 = p_i$ and that the eigenvalues of $\mathbf{K}\mathbf{P}_0$ are $w_j = d_j p_j$. \square

Once the eigenvalues of \mathbf{K} are given, the calculation of $\psi_i^{h^2, H^2}$ and $\xi_i^{h^2, H^2}$ is performed in $O(n)$ complexity. This allows for a simple closed-form formula and fast evaluation of the condition for maximality, even for large datasets.

A.5 Construction of confidence intervals

We wish to build a set of confidence intervals with a coverage probability of $1 - \alpha$ (e.g., 95%). The distribution of \hat{h}^2 depends solely on h^2 , so we may assume without loss of generality that $\sigma_p^2 = 1, \beta = \mathbf{0}_p$ (see A.3). Our approach to CI construction is based on the duality between hypothesis testing and CIs (for a comprehensive treatment of this subject, see [7]).

Let $T(\hat{h}^2; h_0^2, \alpha)$ be any test for the null hypothesis $H_0 : h^2 = h_0^2$, which has size α (i.e., whose type-I error rate is no more than α), and which employs \hat{h}^2 as its test statistic. Let $A(h_0^2, \alpha)$ be the acceptance region of this test, i.e., $T(\hat{h}^2; h_0^2, \alpha)$ does not reject H_0 iff $\hat{h}^2 \in A(h_0^2, \alpha)$. Since the test is of size α , we know that, under the null, $\Pr_{h_0^2}(\hat{h}^2 \in A(h_0^2, \alpha)) \geq 1 - \alpha$. The CI (or, more generally, confidence set) $C(\hat{h}_{\text{obs}}^2; 1 - \alpha)$ that is dual to the test $T(\hat{h}^2; h_0^2, \alpha)$ simply comprises of all parameter values h^2 not rejected by T when observing the estimate value \hat{h}_{obs}^2 , i.e., $C(\hat{h}_{\text{obs}}^2; 1 - \alpha) = \{h^2 \mid \hat{h}_{\text{obs}}^2 \in A(h^2, \alpha)\}$. From this definition, we have that

$$\Pr_{h^2}(h^2 \in C(\hat{h}_{\text{obs}}^2; 1 - \alpha)) = \Pr_{h^2}(\hat{h}_{\text{obs}}^2 \in A(h^2, \alpha)) \geq 1 - \alpha,$$

namely, C is a CI with coverage probability at least $1 - \alpha$. Moreover, if the acceptance region has null probability $1 - \alpha$ exactly (i.e., its associated test is not conservative), then the CI achieves coverage $1 - \alpha$ accurately (otherwise, the CI is also conservative). In the following, we simplify the notation for the CI to $C(\hat{h}^2)$, leaving the dependence on α implicit. We describe a method to construct accurate acceptance regions, and therefore confidence intervals, with coverage probability of $1 - \alpha$.

A.5.1 Preliminaries and Assumptions

Define:

$$p_0(h^2) = \Pr_{h^2}(\hat{h}^2 = 0)$$

$$p_l(h^2) = \Pr_{h^2}(0 < \hat{h}^2 \leq h^2)$$

$$p_r(h^2) = \Pr_{h^2}(h^2 < \hat{h}^2 < 1)$$

$$p_1(h^2) = \Pr_{h^2}(\hat{h}^2 = 1).$$

To prove the accuracy of the CIs, we assume several natural properties of the distribution of the estimators:

Assumption 1. *The boundary probabilities $p_0(h^2)$ and $p_1(h^2)$ are monotone decreasing and increasing functions of h^2 , respectively.*

Figure 4 in the main text indicates that Assumption 1 holds in practice.

Assumption 2. *The required confidence level is larger than the maximum of the boundary probabilities; namely, $1 - \alpha > \max(p_0(0), p_1(1))$.*

To see why Assumption 2 is reasonable, note that the largest boundary probabilities $p_0(0), p_1(1)$ are expected to be ~ 0.5 asymptotically [8]. Indeed, similar values of $\sim 0.5 - 0.6$ are also seen in practice (see Figure 4 in the main text). Meanwhile, $1 - \alpha$ is the required confidence level, with 0.95 being a typical value.

For every $h^2 \in [0, 1]$, the open interval $(0, 1)$ is mapped bijectively to $(p_0(h^2), 1 - p_1(h^2))$

by the CDF F_{h^2} . The quantile function c_β of \hat{h}^2 is:

$$c_\beta(h^2) = \begin{cases} 0 & \text{if } \beta \in [0, p_0(h^2)] \\ F_{h^2}^{-1}(\beta) & \text{if } \beta \in (p_0(h^2), 1 - p_1(h^2)) \\ 1 & \text{if } \beta \in [1 - p_1(h^2), 1]. \end{cases}$$

Note that since estimator distributions are discontinuous at the boundaries, it may be the case that $c_\beta(h^2)$ does not obey $\Pr_{h^2}([0, c_\beta(h^2)]) = \beta$.

Assumption 3. For every β , $c_\beta(h^2)$ is non-decreasing in h^2 .

It follows that

$$\Pr_{h^2}([0, c_\beta(h^2)]) = \begin{cases} p_0(h^2) & \text{if } \beta \in [0, p_0(h^2)] \\ \beta & \text{if } \beta \in (p_0(h^2), 1 - p_1(h^2)) \\ 1 & \text{if } \beta \in [1 - p_1(h^2), 1]. \end{cases}$$

Assumption 4. For all $h^2 \in [0, 1]$, $h^2 \leq c_{1-\alpha}(h^2)$ and $h^2 \geq c_\alpha(h^2)$.

While this may not hold if $1 - \alpha < p_0(h^2)$, it holds in practice for all common values of $1 - \alpha$ (e.g., 95%). Note that since by definition, $c_{1-\alpha}(h^2) \leq c_{1-\alpha/2}(h^2)$ and $c_\alpha(h^2) \geq c_{\alpha/2}(h^2)$, it follows that $h^2 \leq c_{1-\alpha/2}(h^2)$ and $h^2 \geq c_{\alpha/2}(h^2)$.

A.5.2 Acceptance regions

A natural requirement of confidence intervals is that their lower and upper bounds be monotone increasing functions of \hat{h}^2 . Equivalently, we require the lower and upper bounds of the acceptance regions to be monotone increasing functions of h^2 . Therefore, acceptance

regions for increasing values of h^2 are naturally divided into three types: Regions with lower bound 0; then, as the lower bound increases, possibly regions with both upper and lower bounds between 0 and 1; and finally, regions with an upper bound of 1. Specifically, we define the three types of regions to be:

$$\begin{aligned} A_{h^2}^1 &= [0, c_{1-\alpha}(h^2)] \\ A_{h^2}^2 &= [c_{\alpha/2}(h^2), c_{1-\alpha/2}(h^2)] \\ A_{h^2}^3 &= [c_{\alpha}(h^2), 1]. \end{aligned}$$

The three types are defined so that $\Pr_{h^2}(A_{h^2}^i) \geq 1 - \alpha$. It is possible that equality is not obtained. For example, if $p_1(h^2) > \alpha$ at a point h^2 , then $c_{1-\alpha}(h^2) = 1$, and then $\Pr_{h^2}(A_{h^2}^1) = 1$.

A.5.3 Choosing regions

We now define how to choose a region A_{h^2} for each value of h^2 from these three types. There are three requirements from this choice of regions that are desirable:

1. For the respective CIs to be accurate, $\Pr_{h^2}(A_{h^2})$ must be equal to $1 - \alpha$.
2. The true parameter value, h^2 , must be included in each region A_{h^2} .
3. As previously mentioned, the lower and upper bounds of the regions must remain monotone functions of h^2 .

By construction, moving from smaller values of h^2 to larger ones, we consecutively use regions of type I, possibly of type II, and then type III. By definition, we choose a type I region

for $h^2 = 0$ and type III for $h^2 = 1$. There are three scenarios we might encounter:

1. Type II is used. Suppose there is a range of values $h^2 \in [s, t]$ for which $\Pr_{h^2}(A_{h^2}^2) = 1 - \alpha$. Then, we pick s and t as the transition points, i.e.,

$$A_{h^2} = \begin{cases} A_{h^2}^1 = [0, c_{1-\alpha}(h^2)] & \text{if } h^2 \in [0, s) \\ A_{h^2}^2 = [c_{\alpha/2}(h^2), c_{1-\alpha/2}(h^2)] & \text{if } h^2 \in [s, t] \\ A_{h^2}^3 = [c_{\alpha}(h^2), 1] & \text{if } h^2 \in (t, 1] \end{cases}$$

This choice of transitions maintains the monotonicity of the lower and upper bounds of the regions. To see this, note that within each of the parameter ranges, $[0, s)$, (s, t) and $(t, 1]$, monotonicity is obeyed via Assumption 3. At the transition points, monotonicity is maintained by definition of c_{β} , with $c_{1-\alpha}(h^2) \leq c_{1-\alpha/2}(h^2)$ and $c_{\alpha/2}(h^2) \leq c_{\alpha}(h^2)$. Also, from Assumption 4, $h^2 \in A_{h^2}^i$ for $i = 1, 2, 3$. Therefore, $h^2 \in A_{h^2}$ holds for all h^2 . The following proposition shows that the CI is accurate.

Proposition 5. $\Pr_{h^2}(A_{h^2}) = 1 - \alpha$ for all h^2 (i.e. the CI is accurate).

Proof. If $\beta > p_0(0)$, then since p_0 is monotone decreasing (Assumption 1), $\beta > p_0(h^2)$ for all h^2 . Therefore, for such values of β , $\Pr_{h^2}([0, c_{\beta}(h^2)])$ obtains only the values β (when $\beta < 1 - p_1(h^2)$, or equivalently, when $h^2 < p_1^{-1}(1 - \beta)$) and 1 (when $\beta \geq 1 - p_1(h^2)$, or equivalently, when $h^2 \geq p_1^{-1}(1 - \beta)$).

Since $1 - \alpha > p_0(0)$ (Assumption 2), $\Pr_{h^2}([0, c_{1-\alpha}(h^2)]) = 1 - \alpha$ iff $h^2 < p_1^{-1}(\alpha)$ and 1 otherwise. Also, since $1 - \alpha/2 > 1 - \alpha > p_0(0)$, similarly $\Pr_{h^2}([0, c_{1-\alpha/2}(h^2)]) = 1 - \alpha/2$ iff $h^2 < p_1^{-1}(\alpha/2)$ and 1 otherwise.

Since in the range $h^2 \in [s, t]$, $\Pr_{h^2}([0, c_{1-\alpha/2}(h^2)]) = 1 - \alpha/2$, it follows from the above

that $t \leq p_1^{-1}(\alpha/2)$. Since p_1 is monotone increasing (Assumption 1), it follows that $t \leq p_1^{-1}(\alpha/2) \leq p_1^{-1}(\alpha)$. Therefore, in the range $[0, s) \subset [0, t)$, the type I region $A_{h^2}^1$ has the required probability, $Pr_{h^2}(A_{h^2}^1) = 1 - \alpha$.

A similar argument holds for the other direction. If $1 - \beta > p_1(1)$, and since p_1 is monotone increasing, $\beta < 1 - p_1(h^2)$ for all h^2 . Therefore, $Pr_{h^2}([0, c_\beta(h^2)])$ equals $p_0(h^2)$ when $h^2 < p_0^{-1}(\beta)$ and β when $h^2 \geq p_0^{-1}(\beta)$.

Since in the range $h^2 \in [s, t]$, $Pr_{h^2}([0, c_{\alpha/2}(h^2)]) = \alpha/2$, then $s \geq p_0^{-1}(\alpha/2)$. Since p_0 is monotone decreasing, $p_0^{-1}(\alpha) < p_0^{-1}(\alpha/2) \leq s < t$. Therefore, in the range $(t, 1]$, $A_{h^2}^3$ has the required probability, $Pr_{h^2}(A_{h^2}^3) = 1 - \alpha$. Consequentially, $Pr_{h^2}(A_{h^2}) = 1 - \alpha$ for all h^2 . \square

2. Type II is not used, but types I and III achieve accuracy. If there is no range where $Pr_{h^2}(A_{h^2}^2) = 1 - \alpha$, we turn to see if regions of either types I and III may have a probability of $1 - \alpha$ for all $h^2 \in [0, 1]$. We have seen that $Pr_{h^2}(A_{h^2}^1) = Pr_{h^2}([0, c_{1-\alpha}(h^2)]) = 1 - \alpha$ iff $h^2 < p_1^{-1}(\alpha)$ and 1 otherwise. We have also shown that $Pr_{h^2}([0, c_\alpha(h^2)])$ equals $p_0(h^2)$ when $h^2 < p_0^{-1}(\alpha)$ and α when $h^2 \geq p_0^{-1}(\alpha)$. It follows that $Pr_{h^2}(A_{h^2}^3) = Pr_{h^2}([c_\alpha(h^2), 1]) = 1$ when $h^2 < p_0^{-1}(\alpha)$ and $1 - \alpha$ when $h^2 \geq p_0^{-1}(\alpha)$.

Therefore, if $p_0^{-1}(\alpha) < p_1^{-1}(\alpha)$, there does not exist a value h^2 for which neither $A_{h^2}^1$ nor $A_{h^2}^3$ have the required probability of $1 - \alpha$. We pick their average $\delta = (p_0^{-1}(\alpha) + p_1^{-1}(\alpha))/2$ as the transition point, i.e.,

$$A_{h^2} = \begin{cases} A_{h^2}^1 = [0, c_{1-\alpha}(h^2)] & \text{if } h^2 \in [0, \delta) \\ A_{h^2}^3 = [c_\alpha(h^2), 1] & \text{if } h^2 \in [\delta, 1] \end{cases}$$

Consequently, $\Pr_{h^2}(A_{h^2}) = 1 - \alpha$ for all h^2 . Monotonicity of region bounds is trivial in this case, and again, from Assumption 4, $h^2 \in A_{h^2}$ holds for all h^2 .

3. Type II is not used, and types I and III do not achieve accuracy. Finally, if there is no range where $\Pr_{h^2}(A_{h^2}^2) = 1 - \alpha$, and $p_1^{-1}(\alpha) \leq p_0^{-1}(\alpha)$, then in the range $h^2 \in [p_1^{-1}(\alpha), p_0^{-1}(\alpha)]$, $\Pr_{h^2}(A_{h^2}^2) = 1$ for all three region types. In this case, we use randomized confidence intervals [9] to achieve the exact required accuracy.

In this case, it is simpler to discuss the confidence intervals directly. Let

$$C_{H^2} = \begin{cases} [0, U_0] & \text{if } H^2 = 0 \\ [\arg \min_{h^2} \{h^2 | H^2 \in A_{h^2}^1\}, \arg \max_{h^2} \{h^2 | H^2 \in A_{h^2}^3\}] & \text{if } H^2 \in (0, 1) \\ [L_1, 1] & \text{if } H^2 = 1 \end{cases}$$

where U_0 and L_1 are random variables defining the upper bound of C_0 and the lower bound of C_1 , respectively. Let $1 - u(h^2)$ be the cdf of U_0 , and let $l(h^2)$ be the cdf of L_1 . It remains to choose u, l that will be appropriately monotone and will give the required coverage. The total coverage should obey:

$$\begin{aligned} 1 - \alpha &= 1 - p_0(h^2)(1 - u(h^2)) - p_1(h^2)(1 - l(h^2)) \Rightarrow \\ \alpha &= (1 - p_0(h^2))u(h^2) + (1 - p_1(h^2))l(h^2). \end{aligned}$$

The requirement $U_0, L_1 \in [p_1^{-1}(\alpha), p_0^{-1}(\alpha)]$ translates to:

$$u(h^2) = \begin{cases} 1 & \text{if } h^2 \in [0, p_1^{-1}(\alpha)) \\ \text{monotone non-increasing} & \text{if } h^2 \in [p_1^{-1}(\alpha), p_0^{-1}(\alpha)] \\ 0 & \text{if } h^2 \in (p_0^{-1}(\alpha), 1] \end{cases}$$

and

$$l(h^2) = \begin{cases} 0 & \text{if } h^2 \in [0, p_1^{-1}(\alpha)) \\ \text{monotone non-decreasing} & \text{if } h^2 \in [p_1^{-1}(\alpha), p_0^{-1}(\alpha)] \\ 1 & \text{if } h^2 \in (p_0^{-1}(\alpha), 1] \end{cases}$$

Inverting our viewpoint back to acceptance regions, the above is equivalent to choosing $A_{h^2} = A_{h^2}^1$ when $h^2 < p_1^{-1}(\alpha)$ and $A_{h^2} = A_{h^2}^3$ when $h^2 > p_0^{-1}(\alpha)$. When $h^2 \in [p_1^{-1}(\alpha), p_0^{-1}(\alpha)]$, $A_{h^2} = (0, 1) \cup \{0, 1\}$, with the boundary points 0 and 1 included with probabilities $u(h^2)$ and $l(h^2)$ respectively.

When $l(h^2) = 1$,

$$\begin{aligned} \alpha &= (1 - p_0(h^2))u(h^2) + (1 - p_1(h^2)) \\ u(h^2) &= \frac{1 - \alpha - (1 - p_0(h^2))}{p_0(h^2)} = 1 - \frac{\alpha}{p_0(h^2)}. \end{aligned}$$

Similarly, when $u(h^2) = 1$, $l(h^2) = 1 - \frac{\alpha}{p_1(h^2)}$.

The particular form we choose for u and l is

$$u(h^2) = \begin{cases} 1 & \text{if } h^2 \in [0, \delta) \\ 1 - \frac{\alpha}{p_0(h^2)} & \text{if } h^2 \in [\delta, p_0^{-1}(\alpha)] \\ 0 & \text{if } h^2 \in (p_0^{-1}(\alpha), 1] \end{cases}$$

and

$$l(h^2) = \begin{cases} 0 & \text{if } h^2 \in [0, p_1^{-1}(\alpha)) \\ 1 - \frac{\alpha}{p_1(h^2)} & \text{if } h^2 \in [p_1^{-1}(\alpha), \delta] \\ 1 & \text{if } h^2 \in (\delta, 1] \end{cases}$$

where $\delta = (p_0^{-1}(\alpha) + p_1^{-1}(\alpha))/2$, as before.

A.6 Variance of estimators

The main method of calculating the variance of the estimator, applied by all widely used LMM methods, employs the Fisher information matrix, or a variant of which, possibly applying the delta method in addition [10]. The (expected) Fisher information matrix of an estimator $\hat{\boldsymbol{\theta}}$ is the matrix $\mathcal{I}(\boldsymbol{\theta})_{i,j} = \mathbb{E} \left[-\frac{\partial}{\partial \theta_i \partial \theta_j} \ell(\boldsymbol{\theta}; \mathbf{y}) \right]$. Asymptotically, under certain regularity conditions, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta})^{-1})$. According to the delta method, the asymptotic distribution of a function $f(\boldsymbol{\theta})$ satisfies $\sqrt{n}(f(\hat{\boldsymbol{\theta}}) - f(\boldsymbol{\theta})) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \nabla f(\boldsymbol{\theta})^T \mathcal{I}(\boldsymbol{\theta})^{-1} \nabla f(\boldsymbol{\theta}))$.

GCTA uses the Average Information [11] (AI) matrix to calculate the variance of σ_g^2 and σ_e^2 . The AI matrix is an average between the expected and the observed Fisher information

matrices. For the REML method, this is the matrix:

$$\mathcal{J} = \frac{1}{2} \cdot \begin{pmatrix} \mathbf{y}^T \mathbf{Q} \mathbf{K} \mathbf{Q} \mathbf{K} \mathbf{Q} \mathbf{y} & \mathbf{y}^T \mathbf{Q} \mathbf{K} \mathbf{Q} \mathbf{Q} \mathbf{y} \\ \mathbf{y}^T \mathbf{Q} \mathbf{Q} \mathbf{K} \mathbf{Q} \mathbf{y} & \mathbf{y}^T \mathbf{Q} \mathbf{Q} \mathbf{Q} \mathbf{y} \end{pmatrix},$$

where $\mathbf{Q} = \Sigma^{-1} - \Sigma^{-1} \mathbf{X} (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1}$, with $\Sigma = \sigma_g^2 \mathbf{K} + \sigma_e^2 \mathbf{I}$. Then, the delta method is used to calculate the variance of \hat{h}^2 :

$$\text{Var}(\hat{h}^2) = (\hat{\sigma}_g^2 + \hat{\sigma}_e^2)^{-4} \begin{pmatrix} \hat{\sigma}_e^2 & -\hat{\sigma}_g^2 \end{pmatrix} \mathcal{J}^{-1} |_{\sigma_g^2 = \hat{\sigma}_g^2, \sigma_e^2 = \hat{\sigma}_e^2} \begin{pmatrix} \hat{\sigma}_e^2 \\ -\hat{\sigma}_g^2 \end{pmatrix}.$$

GEMMA instead applies the delta method on a closed-form formula of the ratio σ_g^2/σ_e^2 [12].

Figures and Tables

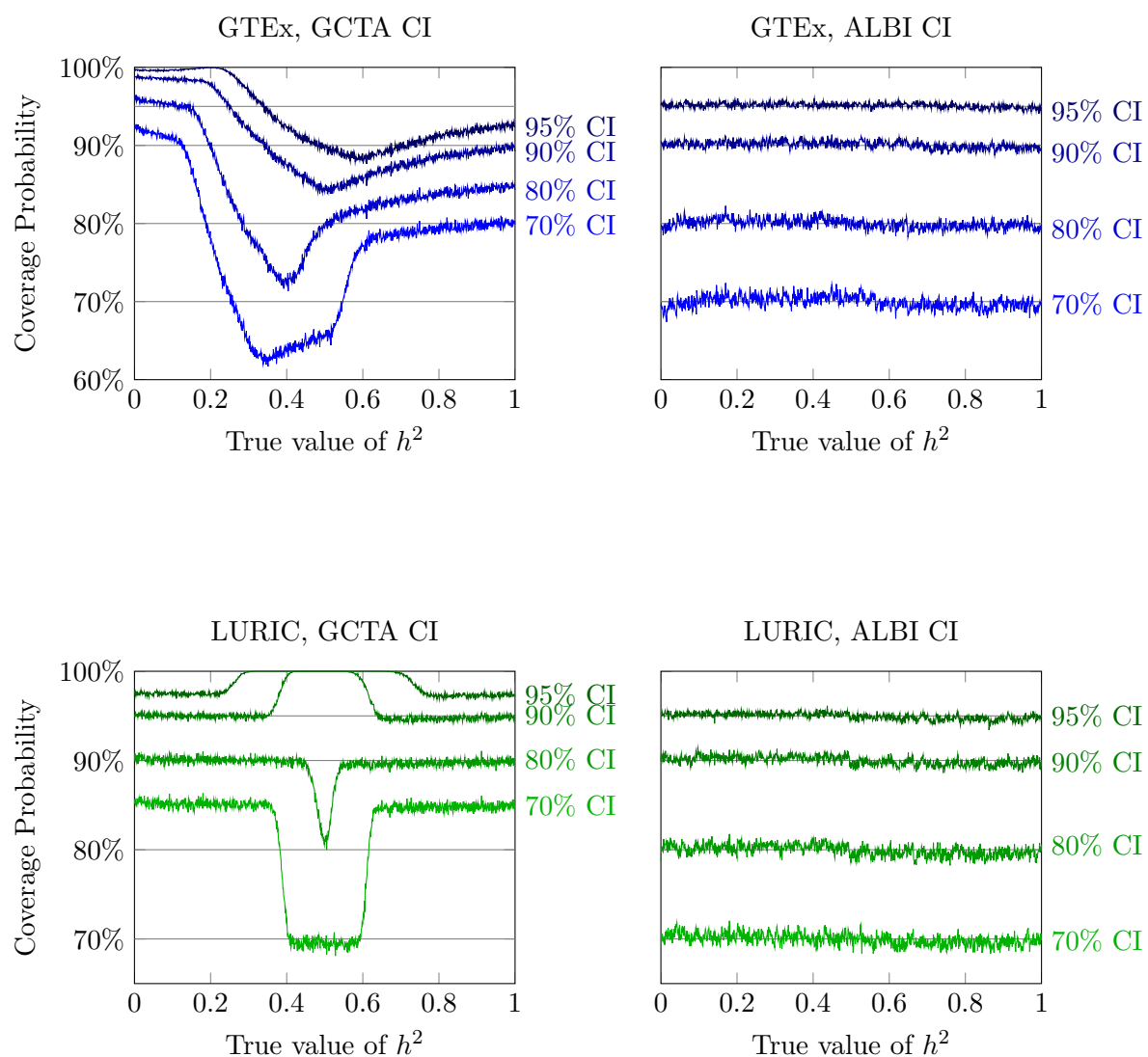


Figure 1 Accuracy of CIs. The left panes depict the true coverage probability of GCTA's confidence intervals (CIs), based on the normal approximation, on the GTEx and LURIC datasets. The right panes depict the coverage probabilities of the ALBI CIs. The coverage probabilities are shown for CIs designed to have coverage probabilities of 70%, 80%, 90% and 95%. GCTA's CIs are often far from the correct confidence level, while ALBI's achieve accurate coverage.

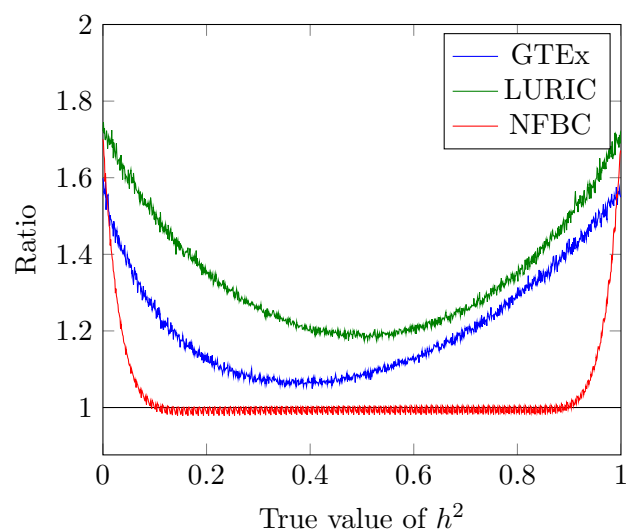


Figure 2 Accuracy of standard errors. The ratio between the mean standard error derived from GCTA, and the empirical standard deviation of the REML estimator \hat{h}^2 , as a function of the true h^2 , for the studied datasets. The discrepancy is high, especially for GTE_x and LURIC, with ratios up to $\times 1.7$.

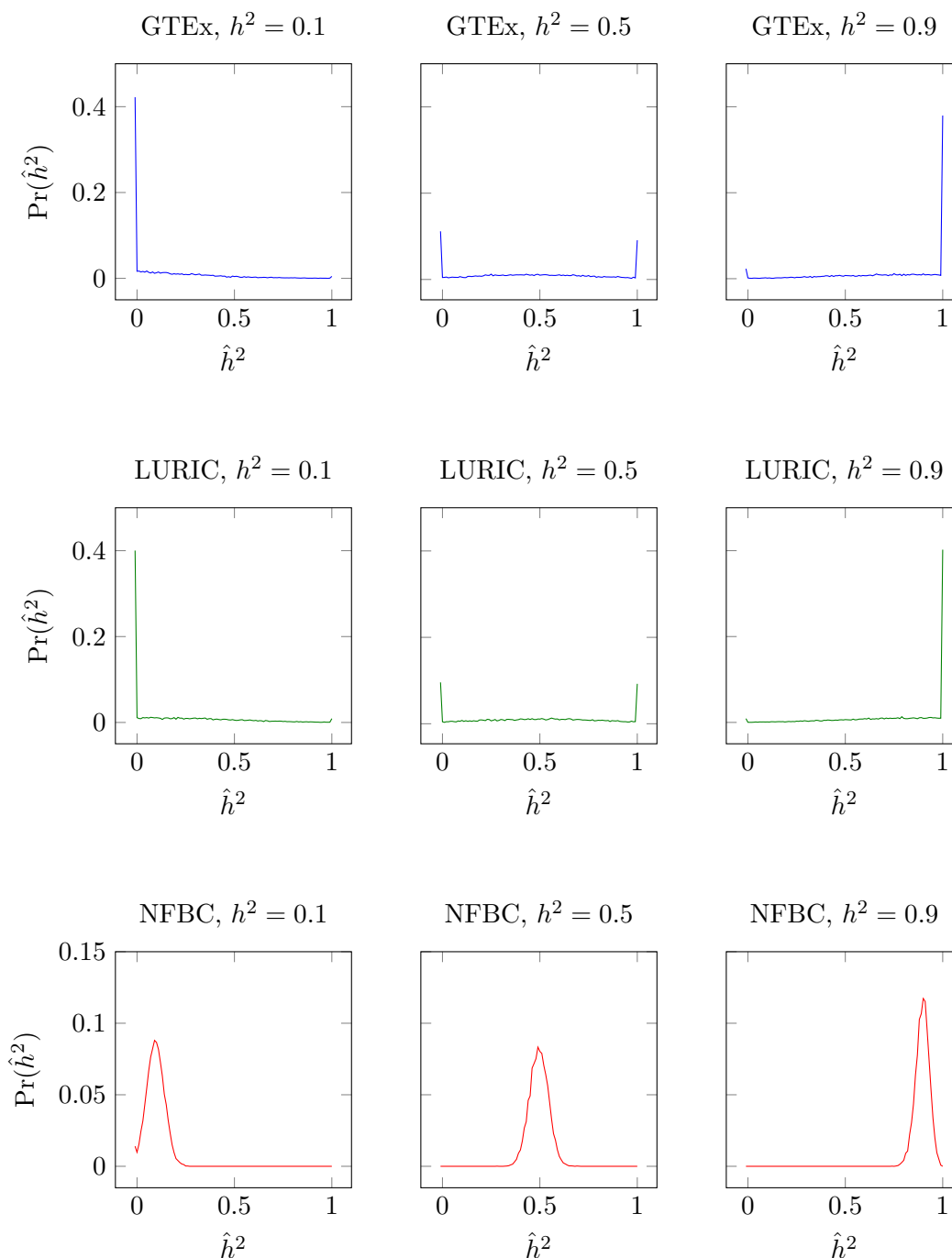


Figure 3 Distributions of the heritability estimator. The density of \hat{h}^2 for $h^2 = 0.1, 0.5, 0.9$ in the studied datasets, under the LMM. Since the distribution of \hat{h}^2 is continuous in the open interval $(0, 1)$, we divide this range into equally sized bins of a fixed size (0.01), and instead estimate the probability mass function of a random variable taking values in the set $\{0, (0, 0.01], (0.01, 0.02], \dots, (0.09, 1), 1\}$. Estimator distributions are often far from being normal, hence the normal approximation seems highly questionable.

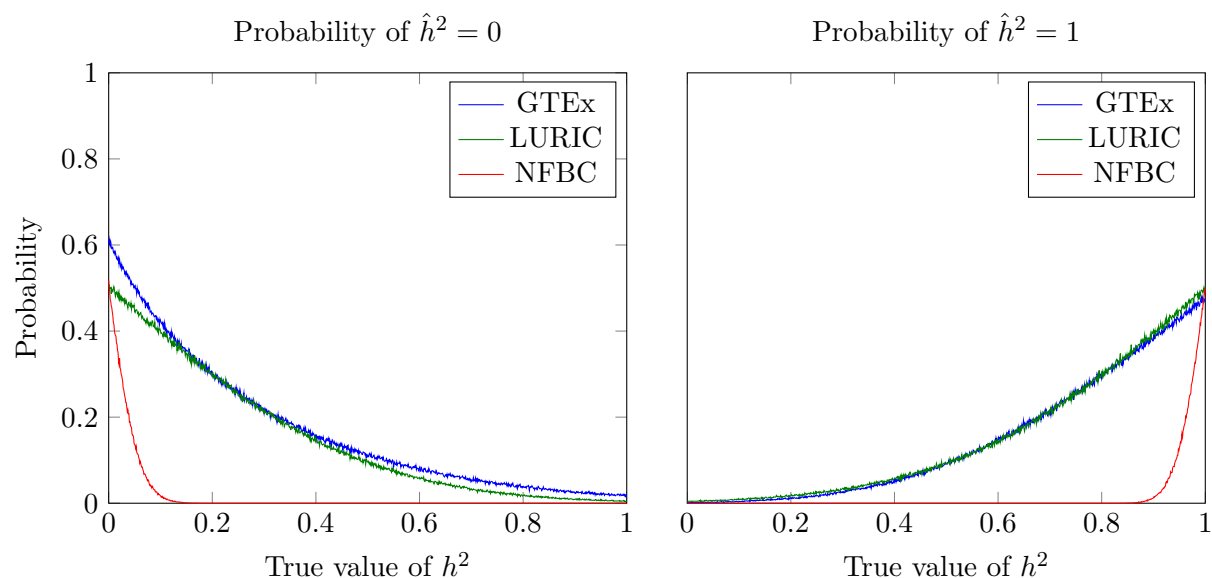


Figure 4 Probability of boundary heritability estimates. The probability of estimating $\hat{h}^2 = 0$ or 1, for all possible values of true underlying heritability values h^2 , for the studied datasets. For GTEEx and LURIC, the probability of $\hat{h}^2 = 0$ (1) is high, especially for small (large) values. For NFBC, this phenomenon is limited.

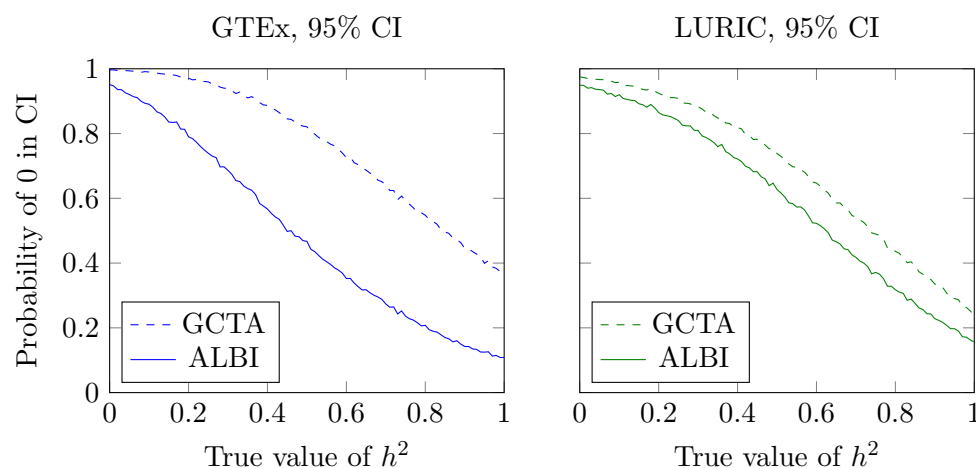


Figure 5 Probability of zero heritability in CI. The probability of $h^2 = 0$ being included in the CI, as a function of the true value of h^2 , for the GTEX and LURIC datasets. These probabilities are shown for GCTA's CI and ALBI's CI, designed to have a confidence level of 95%. It can be seen that CIs derived from the normal approximation tend to include $h^2 = 0$ more than necessary.

Dataset	ALBI (0.01)	ALBI (0.001)	pylmm	GCTA
GTEX, 100 random samples	0.04 sec	0.27 sec	3.94 sec	1.05 min
LURIC, 100 random samples	0.12 sec	1.27 sec	5.7 sec	3.98 min
NFBC, 100 random samples	0.9 sec	7.93 sec	1.31 min	> 8 hours
GTEX, 1000 random samples	0.27 sec	1.9 sec	37.9 sec	4.68 min
LURIC, 1000 random samples	0.73 sec	8.85 sec	54.15 sec	54.6 min
NFBC, 1000 random samples	7.42 sec	1.25 min	6.06 min	> 1 day

Table 1 Benchmarks. Running times of ALBI vs. other brute-force methods. We compare the computational costs of ALBI to that of pylmm [13] and GCTA [1] (see Methods). We generated the estimator distribution for $h^2 = 0.5$. For ALBI, we estimated the distributions at a precision of 0.01 and 0.001, using 100 and 1,000 random samples. For pylmm and GCTA, we generated 100 and 1,000 random phenotypes. GCTA did not converge for ~7% of the random samples. Running times are reported for the datasets of GTEX (185 individuals), LURIC (867 individuals) and NFBC (5,236 individuals).

Supplementary Figures and Tables

REML estimator bias

We evaluated the empirical bias of the REML estimator \hat{h}^2 as a function of h^2 . It is evident that the REML estimator may be biased (Figure S1). The bias is especially evident for low and high true values of h^2 . This is also true for the estimator of σ_p^2 (not shown).

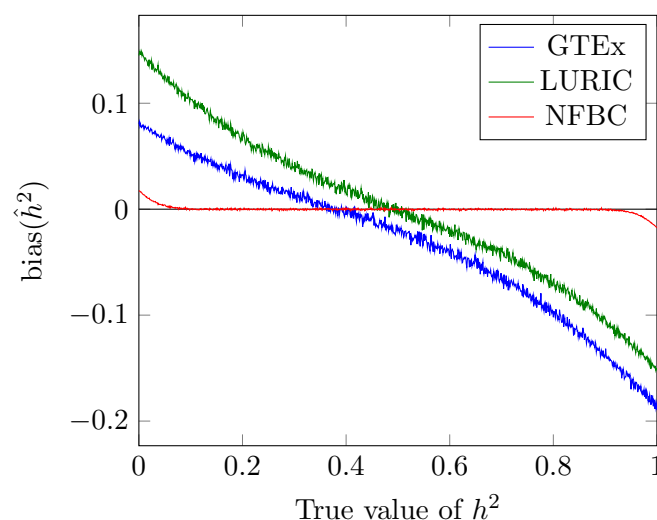


Figure S1 Bias of the heritability estimator. The bias of \hat{h}^2 (defined as $E(\hat{h}^2) - h^2$), for all possible values of true underlying heritability h^2 , for the studied datasets. The GTEx and LURIC datasets create highly biased estimators, while for the NFBC the estimator is mostly unbiased.

References

1. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76–82 (2011).

2. Hartley, H. O. & Rao, J. N. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108 (1967).
3. Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).
4. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. **8** (2011).
5. Harville, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320–338 (1977).
6. Crainiceanu, C., Ruppert, D. & Vogelsang, T. Probability that the MLE of a variance component is zero with applications to likelihood ratio tests. *Unpublished manuscript* (2002).
7. Casella, G. & Berger, R. L. *Statistical inference* (Duxbury Pacific Grove, CA, 2002).
8. Self, S. G. & Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610 (1987).
9. Geyer, C. J. & Meeden, G. D. Fuzzy and randomized confidence intervals and P-values. *Statistical Science*, 358–366 (2005).
10. Wasserman, L. *All of statistics: a concise course in statistical inference* (Springer Science & Business Media, 2013).
11. Gilmour, A. R., Thompson, R. & Cullis, B. R. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, 1440–1450 (1995).
12. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**, 821–4 (2012).

13. Furlotte, N. A. & Eskin, E. Efficient multiple trait association and estimation of genetic correlation using the matrix-variate linear mixed-model. *Genetics* **200**, 59–68 (2015).