Using Stochastic Approximation Techniques to Efficiently Construct Confidence Intervals for Heritability

Regev Schweiger[1], Eyal Fisher[2], Elior Rahmani[1], Liat Shenhav[2], Saharon Rosset[2], and Eran Halperin[3,4]

[1] Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel, schweiger@post.tau.ac.il
[2] School of Mathematical Sciences, Department of Statistics, Tel Aviv University, Tel Aviv, Israel
[3] Department of Computer Science, University of California, Los Angeles, CA, USA
[4] Department of Anesthesiology and Perioperative Medicine, University of California, Los Angeles, CA, USA

Abstract. Estimation of heritability is an important task in genetics. The use of linear mixed models (LMMs) to determine narrow-sense SNP-heritability and related quantities has received much recent attention, due of its ability to account for variants with small effect sizes. Typically, heritability estimation under LMMs uses the restricted maximum likelihood (REML) approach. The common way to report the uncertainty in REML estimation uses standard errors (SE), which rely on asymptotic properties. However, these assumptions are often violated because of the bounded parameter space, statistical dependencies, and limited sample size, leading to biased estimates and inflated or deflated confidence intervals. In addition, for larger datasets (e.g., tens of thousands of individuals), the construction of SEs itself may require considerable time, as it requires expensive matrix inversions and multiplications.

Here, we present FIESTA (Fast confidence IntErvals using STochastic Approximation), a method for constructing accurate confidence intervals (CIs). FIESTA is based on parametric bootstrap sampling, and therefore avoids unjustified assumptions on the distribution of the heritability estimator. FIESTA uses stochastic approximation techniques, which accelerate the construction of CIs by several orders of magnitude, compared to previous approaches as well as to the analytical approximation used by SEs. FIESTA builds accurate CIs rapidly, e.g., requiring only several seconds for datasets of tens of thousands of individuals, making FIESTA a very fast solution to the problem of building accurate CIs for heritability for all dataset sizes.

1 Introduction

Heritability, or the proportion of phenotypic variation that is explained by genetic variation, is an important population parameter in human genetics, in

evolution, in plant and animal breeding, and more. Estimating the heritability has been traditionally performed using related individuals such as in twin studies or pedigree designs [1–3]. More recently, genetic variation has been estimated using genetic marker information, and in particular in genome-wide association studies (GWAS) [4, 5], which have identified thousands of genetic variants that are associated with dozens of common diseases. However, genome-wide significant associations were generally found to explain only a small proportion of the heritability of complex diseases.

To cope with this challenge, linear mixed model (LMM) approaches [6–13] have been applied to estimate the heritability explained by common SNPs (the narrow-sense SNP-heritability, to which we refer as heritability, and denote by h^2) from cohorts of unrelated individuals, such as those found in GWAS [14]. Estimation under the LMM is usually performed using restricted maximum likelihood (REML) estimation, and is implemented in some widely used tools, like the GCTA software package [15]. LMMs utilize all variants from a GWAS, and not just the variants that are statistically significant, and therefore is able to account for variants with small effect sizes.

As in any statistical analysis, the process of estimating the heritability suffers from statistical uncertainty. Typically, confidence intervals (CIs) are reported alongside with point estimates to quantify this uncertainty. Usually, such CIs are constructed from standard errors (SEs), which make the assumption that the estimators asymptotically follow a normal distribution. However, it has been shown [13,16–20] that such CIs can be highly inaccurate. This is because estimators do not necessarily obey the conditions required for them to asymptotically follow the normal distribution. 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.

To handle these issues, previous approaches have taken several directions. Non-standard asymptotic theory for boundary and near-boundary maximum likelihood estimates has been developed (e.g., [21–23]), and it has been suggested to replace the asymptotic normality assumption with the asymptotics developed for the non-standard boundary case [24]. Visscher et al. [25] derived an analytical expression for the asymptotic variance of the heritability estimator in a range of pedigree- and marker-based experimental designs. Unfortunately, these conditions typically do not hold for genomic datasets, mainly due to the limited sample size, making either of these approximations ineffective [20]. Other approaches include hierarchical bootstrapping schemes, e.g., [26]; extending the REML estimation method with Bayesian priors, e.g., [27, 28]; using alternative statistics as a basis for building CIs [17, 29, 30]; or using Bayesian posterior distribution of the heritability value [31].

An alternative approach is the parametric bootstrap test inversion technique, which constructs CIs via sampling phenotypes, performing heritability estimation on the sampled phenotypes, estimating the distribution of the heritability estimator and using these estimates as a basis for CI construction [32]. The main

advantage of using a parametric bootstrap approach is that it does not require any assumptions on the distribution of the heritability estimator or of Bayesian priors. As a naïve implementation of this approach would be computationally prohibitive, the ALBI method [20] utilizes a highly accurate approximation that allows an efficient construction of accurate CIs. However, ALBI still requires a preprocessing step. Newer datasets (e.g. the UK Biobank [33]) may contain tens or hundreds of thousands of individuals, for which this step may require hours of computation time. In addition, the need for a preprocessing step can be an obstacle in the adoption of a better CI construction method.

In this paper, we introduce FIESTA (Fast confidence IntErvals using STochastic Approximation), which dramatically reduces the running time of CI construction by several orders of magnitude, e.g., to mere seconds for dataset with tens of thousands of individuals, compared to hours or days. The key ingredient of our approach is a CI construction algorithm from the field of stochastic approximation (for a review, see [34]). Originating in the work of Robbins and Monro [35], stochastic approximation algorithms are recursive update rules that can be used, among other things, to solve optimization problems or function inversion problems when the collected data is subject to noise. It has been shown [36] that stochastic approximation can be used to construct CIs for general families of parametric distributions, given the ability to randomly sample from them, and this is the approach we employ here. We validate FIESTA on two real datasets, the Northern Finland Birth Cohort (NFBC) dataset [37] and the Wellcome Trust Case Control Consortium 2 (WTCCC2) [38] dataset.

In addition to the significant speedup in time, FIESTA requires no preprocessing step beyond calculating the eigendecomposition of the kinship matrix, which is usually already performed as a part of heritability estimation. Finally, we show that FIESTA is even significantly faster than the analytical SE formulation. In summary, FIESTA can effectively be used extremely easily to rapidly generate accurate CIs for REML heritability estimates. FIESTA is available as part of the ALBI toolkit at `https://github.com/cozygene/albi`.

2 Results

2.1 A Faster Method for Calculating CIs for Heritability

CIs constructed from standard errors, which are based on the assumption of a normal distribution for the heritability estimators, were previously shown to be inaccurate [13, 16–20]. In this paper, we introduce FIESTA, a method that generates accurate CIs for h^2, the true heritability value, given $\hat{h}^2$, the restricted maximum likelihood (REML) estimator for h^2 (see Methods). FIESTA uses the principle of test inversion to construct accurate CIs, using a stochastic approximation method that directly estimates the CI boundaries. We review FIESTA below; for a full description, see Methods.

The methodology of test inversion can be described as follows. The estimator $\hat{h}^2$ is a function of the phenotype, which is a random variable whose distribution

depends on h^2, assuming a fixed kinship matrix. Therefore, $\hat{h}^2$ is distributed differently for every value of h^2. For each true value of h^2, we select a subset of possible $\hat{h}^2$ values that has a sampling probability of $1-\alpha$, where $\hat{h}^2$ is distributed under the assumption of a true heritability value h^2. We define this subset to be the acceptance region for that value of h^2. The CI accompanying an estimate $\hat{h}^2$ is the interval containing all values of h^2 whose acceptance region includes $\hat{h}^2$, namely, for which $\hat{h}^2$ does not imply the rejection of the null hypothesis that the true heritability value is h^2, with a significance level of α.

It remains to define suitable acceptance regions. In the Methods section, we review our scheme for defining acceptance regions. A basic ingredient of our construction of acceptance regions is inverting certain quantile functions of the distribution of $\hat{h}^2$, as a function of h^2. For example, finding the inverse of a value H^2 of the 95%-quantile function is finding a heritability value h^2 for which $\Pr_{h^2}(\hat{h}^2 \leq H^2) = 0.95$, i.e., the probability to get an heritability estimate of H^2 or below is precisely 95%, when $\hat{h}^2$ is distributed with the heritability value h^2.

Instead of carrying out this task by a full parametric bootstrap estimate of the distribution of the estimator, we employ a technique from the field of stochastic approximation to achieve the same results with a fraction of the computational cost. The modified Robbins-Monro procedure [39], described in the Methods section, is an iterative method that finds the inverse of the quantile function of a one-parameter distribution. It operates by iteratively (1) drawing a sample with a true heritability value equal to our current guess for the required inverse value, (2) comparing its estimated heritability to H^2; (3) updating our current guess accordingly, by moving in the right direction, with a step size that decreases with the number of iterations. An additional speedup is acquired by using a fast method to calculate the derivative of likelihood of the sample, and using the derivative to compare its estimated heritability to H^2, instead of performing the full likelihood maximization.

We applied FIESTA to construct 95% CIs for the NFBC dataset [37] and the WTCCC2 dataset [38], as seen in Figure 1. We then turned to verify the accuracy of these CIs, which can be measured as follows. Draw multiple phenotype vectors from the distribution assumed by the LMM with parameters that correspond to a true heritability value h^2. From each such phenotype, construct a CI for its estimated heritability with a confidence level of, e.g., 95%. If the constructed CIs are accurate, then they should cover the true underlying h^2 95% of the time. Then, check the percentage of times in which the CI covered h^2, as a function of h^2. We measured the accuracy of FIESTA, with CIs designed to have a coverage of 95%. The results are shown in Figure 2, demonstrating that FIESTA accurately achieves the desired confidence levels.

2.2 Benchmarks

We compared the speed of the stochastic approximation approach, implemented in FIESTA, with that of using the parametric bootstrap for estimating the distribution of heritability estimator. The latter was tested either as implemented naively by using either GCTA [15] and pylmm [40], or by using ALBI [20]. Both

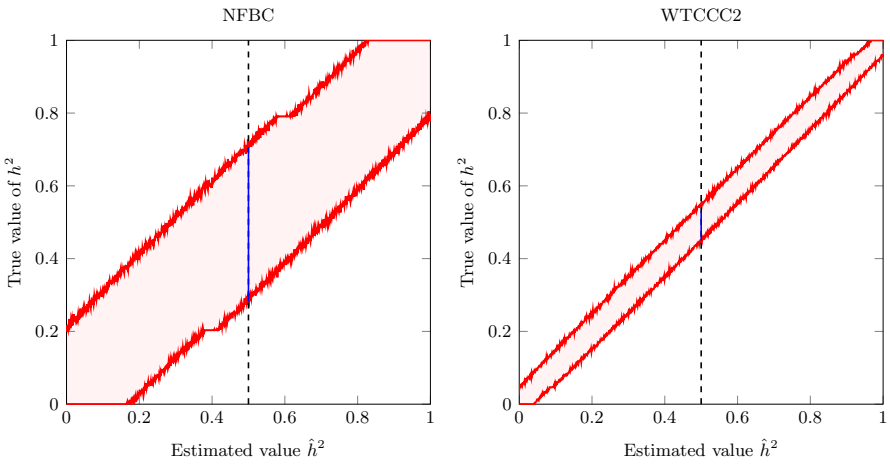

Fig. 1: **95% CIs for the NFBC and WTCCC2 datasets.** Accurate 95% CIs constructed for the NFBC dataset [37] (left) and the WTCCC2 [38] dataset (right) by FIESTA. For each $\hat{h}^2$ on a fine grid of 1000 values (x axis), we constructed a CI, whose boundaries are shown (y axis). For example, for $\hat{h}^2 = 0.5$ (denoted by a dashed line), the CI for NFBC is $[0.282, 0.705]$ (denoted by a full line).

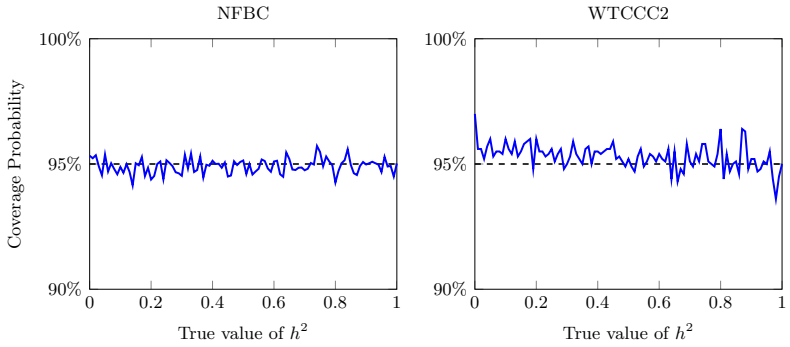

Fig. 2: **Accuracy of CIs for the NFBC and WTCCC2 datasets.** The coverage probabilities of the FIESTA CIs. The coverage probabilities are shown for CIs designed to have coverage probabilities of 95%. The CIs achieve accurate coverage.

the stochastic approximation and parametric bootstrap approaches require the calculation of the eigendecomposition of the kinship matrix. As this is already often a part of the heritability estimation algorithm, its calculation time is separated in the benchmarks. In the Discussion, we discuss how this step could be avoided altogether.

One difference between the two approaches is that the bootstrap approach performs a lengthy preprocessing step that estimates many distributions. Once these distributions are estimated, constructing a CI is very rapid. In contrast, the stochastic approximation approach does not perform a preprocessing step, but performs a non-trivial calculation per CI.

The construction of a single CI with FIESTA consists of calculating six to eight values using the modified Robbins-Monro procedure (see Methods). The first four values depend only on the kinship matrix, but not on the heritability estimate for which we construct a CI, so they need to be calculated only once per kinship matrix, and can then be shared between several CIs. Each modified Robbins-Monro run has the complexity of $O(nT)$, where n is the number of individuals in the sample and T is the number of iterations (in the order of 1,000; see Methods). Therefore, in total, the time complexity to construct K CIs with FIESTA grows linearly with K, T and n.

We also compared FIESTA to the performance of the analytical SE approach. While often inaccurate, analytical SEs are often the go-to method by many practitioners: First, their calculation is conceptually easy to understand, since a closed-form formula exists for the SEs (see Appendix); second, using a closed-form expression is often perceived as faster than more involved algorithmic procedures. However, this is not the case for heritability estimation, as SEs are calculated using variants of the Fisher information matrix (e.g., the AI matrix, as in GCTA [15]), whose calculation requires matrix-by-vector multiplications, which are $O(n^2)$. In contrast, FIESTA is linear in n, giving it an advantage at larger datasets in particular.

We performed a benchmark to evaluate FIESTA, using the NFBC and WTCCC2 datasets. We estimated the distributions of $\hat{h}^2$ for $h^2 = 0, 0.01, \ldots, 1$, with GCTA [15] and pylmm [40], both of which perform full estimation, using 1,000 random bootstrap samples. For the same task, we also used ALBI [20], at a grid resolution of 0.001. The accuracy of CIs constructed according to the full estimation approach, as implemented in ALBI, are shown in the Appendix. As explained above, the time of construction of CIs given these distributions is negligible relative to the time required for the estimation of the distributions. We also constructed analytical SEs for both datasets using the AI method (see Appendix). These times are reported in Table 1.

As a comparison, we used FIESTA to construct varying number of CIs, using 1,000 iterations in the modified Robbins-Monro procedure (see Methods). In Table 1, it can be seen that FIESTA is significantly faster, particularly when few CIs are needed. We also note that FIESTA is currently implemented in the Python language, using the numpy package; a significant additional speedup can be obtained by migrating to a compiled language, e.g., C++.

We then continued to investigate the stability of CI construction and its dependency on the number of iterations. We ran FIESTA 100 times to construct CI for the NFBC and WTCCC2 datasets using 200, 500, 1,000 or 2,000 iterations. We measured the variance in the constructed CI endpoints (Table 2). As expected, the variance decreases with the number of iterations. In addition, we measured the mean and variance of the coverage of CIs under a grid of true heritability values. Here, also, we observed that variance of coverage decreases with the number of iterations. We note that 500 iterations are sufficient for reasonably accurate CIs for these datasets, and that the coverage of even 200 iterations is only slightly biased downwards.

Table 1: **Benchmarks.** Running times of FIESTA, compared with previous methods (see Results for more details). Running times are reported for the NFBC (2,520 individuals) and WTCCC2 (13,950 individuals) datasets.

Algorithm	**Software**	**Time for NFBC**	**Time for WTCCC2**
Eigen-decomp.	GCTA	50 seconds	2 hours
Full bootstrap	GCTA	$>$ 30 days	$>$ 30 days
Full bootstrap	pylmm	3.8 hours	$>$ 8 days
Full bootstrap	ALBI	5.35 minutes	2.5 hours
Analytical SEs	n/a	~3.1 sec × # of CIs, e.g.: 1 CI, ~3 seconds 5 CIs, ~15 seconds 10 CIs, ~31 seconds 50 CIs, ~2.6 minutes	~6.2 min × # of CIs, e.g.: 1 CI, ~6 minutes 5 CIs, ~31 minutes 10 CIs, ~1 hours 50 CIs, ~5 hours
Stochastic Approximation	**FIESTA**	**~1.8 sec + 0.6 sec × # of CIs**, e.g.: 1 CI, ~**3 seconds** 5 CIs, ~**6 seconds** 10 CIs, ~**8 seconds** 50 CIs, ~**33 seconds**	**~6 sec + 2.8 sec × # of CIs**, e.g.: 1 CI, ~**9 seconds** 5 CIs, ~**20 seconds** 10 CIs, ~**34 seconds** 50 CIs, ~**2.4 minutes**

3 Methods

For clarity of presentation, we begin by defining the heritability under the LMM, and briefly reviewing stochastic approximation and its relevance to finding CIs. Finally, we introduce FIESTA, our improved method for faster construction of CIs for heritability.

3.1 The Linear Mixed Model and REML

We consider the following standard linear mixed model (see [41] for a detailed review). Let n be the number of individuals and m is the number of SNPs. Let $\mathbf{y}$ be a $n \times 1$ vector of phenotype measurements for each individual. Let $\mathbf{X}$ be a

Table 2: **Stability of CI construction.** 95% CIs for the NFBC and WTCCC2 datasets were constructed 100 times, with either 200, 500, 1,000 or 2,000 iterations. CIs were constructed for $\hat{h}^2 = 0, 0.001, \ldots, 1$. In order to assess the variance of the construction process, the mean empirical standard error (SE) of the lower and upper endpoints is reported, where the mean was calculated over all non-constant endpoints, across all $\hat{h}^2$ values. In addition, the CI coverage for $h^2 = 0, 0.01, \ldots, 1$ was calculated as in Figure 2. The average mean and SE across all 100 runs, calculated across all h^2, is reported.

Dataset	**NFBC**				**WTCCC2**			
No. of iterations	200	500	1,000	2,000	200	500	1,000	2,000
CI lower point SE	0.0201	0.0132	0.0094	0.0067	0.0050	0.0032	0.0023	0.0016
CI upper point SE	0.0206	0.0133	0.0096	0.0070	0.0050	0.0031	0.0023	0.0016
Mean coverage	94.20%	94.71%	94.87%	94.95%	94.720%	95.217%	95.323%	95.373%
SE of coverage	0.45%	0.34%	0.30%	0.28%	0.781%	0.575%	0.486%	0.442%

$n \times p$ matrix of p covariates (possibly including an intercept vector $\mathbf{1}_n$ as a first column, as well as other covariates such as sex, age, etc.). Let $\mathbf{Z}$ be the $n \times m$ standardized genotype matrix, i.e., columns have zero mean and unit variance. Let $\boldsymbol{\beta}$ be a $p \times 1$ vector of fixed effects, $\mathbf{s}$ a $m \times 1$ vector of random effects, and $\mathbf{e}$ a $n \times 1$ vector of errors. Then, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{s} + \mathbf{e}$. We assume $\mathbf{s}$ and $\mathbf{e}$ are statistically independent and are distributed normally as $\mathbf{s} \sim \mathcal{N}\left(\mathbf{0}_m, \frac{1}{m}\sigma_g^2\mathbf{I}_m\right)$, $\mathbf{e} \sim \mathcal{N}\left(\mathbf{0}_n, \sigma_e^2\mathbf{I}_n\right)$. The fixed effects $\boldsymbol{\beta}$ and the coefficients σ_g^2 and σ_e^2 are the parameters of the model.

Define $\mathbf{K} = \frac{1}{m}\mathbf{Z}\mathbf{Z}^\mathbf{T}$. Typically, $\mathbf{K}$ is commonly called the kinship matrix, or the genetic relationship matrix. Under these conditions, it follows [14] that:

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

The narrow-sense heritability due to genotyped common SNPs is defined as the proportion of total variance explained by genetic factors [42]:

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

Defining $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$, Equation 1 becomes: $\mathbf{y} \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma_p^2\mathbf{V}_{h^2}\right)$, where $\mathbf{V}_{h^2} = h^2\mathbf{K} + (1 - h^2)\mathbf{I}_n$.

The most common way of estimating h^2 is restricted maximum likelihood (REML) estimation. REML consists of maximizing the likelihood function associated with the projection of the phenotype onto the subspace orthongonal to that of the fixed effects of the model [43]. In [20], it is shown 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

$\sigma_p^2 = 1$ and $\boldsymbol{\beta} = \mathbf{0}_p$, which substantially simplifies the problem; namely, we may focus on properties of the distribution $\mathcal{N}(\mathbf{0}_n, \mathbf{V}_{h^2})$ instead of the more general $\mathcal{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma_p^2 \mathbf{V}_{h^2}\right)$.

3.2 Confidence Intervals for h^2

We wish to build confidence intervals with a coverage probability of $1-\alpha$ (e.g., 95%). The full derivation is developed in [20], and is reviewed in the Appendix; we cite the results here.

Let $c_\beta(h^2)$ be the β-th quantile function of $\hat{h}^2$, when the true heritability is h^2; i.e. $\Pr_{h^2}(\hat{h}^2 \leq c_\beta(h^2)) = \beta$. Define s and t to be the values for which $\Pr_{h^2=s}(\hat{h}^2 = 0) = \alpha/2$ and $\Pr_{h^2=t}(\hat{h}^2 = 1) = \alpha/2$. In addition, let $s^* = c_{1-\alpha}(0), t^* = c_\alpha(1)$. Then the lower and upper CI boundaries for an estimate H^2 are given, respectively, by

$$l_{H^2} = \begin{cases} 0 & \text{if } H^2 \leq s^* \\ c_{1-\alpha}^{-1}(H^2) & \text{if } c_{1-\alpha}^{-1}(H^2) < s \\ s & \text{if } s \in [c_{1-\alpha/2}^{-1}(H^2), c_{1-\alpha}^{-1}(H^2)] \\ c_{1-\alpha/2}^{-1}(H^2) & \text{if } s < c_{1-\alpha/2}^{-1}(H^2) \end{cases} \tag{3}$$

and

$$u_{H^2} = \begin{cases} c_{1-\alpha/2}^{-1}(H^2) & \text{if } c_{\alpha/2}^{-1}(H^2) < t \\ t & \text{if } t \in [c_\alpha^{-1}(H^2), c_{\alpha/2}^{-1}(H^2)] \\ c_\alpha^{-1}(H^2) & \text{if } t < c_\alpha^{-1}(H^2) \\ 1 & \text{if } t^* \leq H^2\,. \end{cases} \tag{4}$$

3.3 Using Stochastic Approximation to Calculate CIs

Robbins-Monro. Stochastic approximation methods are a family of iterative stochastic optimization algorithms that attempt to find zeroes, inverses or extrema of functions which cannot be computed directly, but only estimated via noisy observations. The classical Robbins-Monro algorithm presents a methodology for solving a function inversion problem, where the function is the expected value of a parametrized family of distributions. Namely, a function $g(\theta)$ is given, for which we want to find an inverse, i.e., a value $\bar{\theta}$ for which $g(\bar{\theta}) = C$, for some constant C. However, the function g is not directly available to us, but rather we are only able to obtain noisy observations from it. The Robbins-Monro procedure is a modification of Newton's method, where the step sizes are instead an appropriately decreasing sequence. Starting with an initial guess, θ_0, at iteration n we obtain a noisy sample y_n from a distribution whose mean is $g(\theta_n)$, and update our estimate with

$$\theta_{n+1} = \theta_n - \gamma_n \cdot (y_n - C) \tag{5}$$

where $\gamma_n = 1/n$. The Robbins-Monro procedure is shown to converge to the correct solution when: (i) the random variables defining our sampling process at each $g(\theta)$ are uniformly bounded; (ii) $g(\theta)$ is nondecreasing; and (iii) $g'(\bar{\theta})$ exists and is positive [35].

Using Robbins-Monro to calculate CIs. Garthwaite and Buckland [36] have used the Robbins-Monro process for finding the endpoints of CIs, as we will now describe. We discuss the case of one-sided CIs, but the application to two-sided CIs is immediate.

Suppose that $[0, u_{\hat{\theta}})$ is the one-sided $1-\alpha$ CI for θ, when data $\mathbf{y}$ has been observed, with an estimate $\hat{\theta} = \hat{\theta}(\mathbf{y})$. Then, the correct endpoint satisfies

$$\Pr_{\theta=u_{\hat{\theta}}}\left(\hat{\theta} \leq \hat{\theta}(\mathbf{y})\right) = \alpha \tag{6}$$

If we define $g(\theta) = \Pr_{\theta}\left(\hat{\theta} \geq \hat{\theta}(\mathbf{y})\right)$ (to make it nondecreasing), then finding $u_{\hat{\theta}}$ is equivalent to finding the inverse of g at $1-\alpha$. However, under these settings, we do not have direct access to g. Rather, we sample a binary random variable Y_θ, indicating that a sample $\mathbf{y}_\theta$ randomly drawn from $g(\theta)$ has an estimate $\hat{\theta}(\mathbf{y}_\theta)$ larger than $\hat{\theta}(\mathbf{y})$. By definition, $\Pr_\theta(Y_\theta) = \Pr_\theta\left(\hat{\theta}(\mathbf{y}_\theta)) \geq \hat{\theta}(\mathbf{y})\right) = g(\theta)$, so the random sample Y_θ has a mean of $g(\theta)$. Effectively, this formulation allows us to use the Robbins-Monro procedure to invert the quantile function as a function of θ. Full asymptotic efficiency can be achieved by multiplying the step size γ_n by some constant c.

In detail, denote by y_n a random sample from the random variable Y_{θ_n}. The update rule is $\theta_{n+1} = \theta_n - c\gamma_n \cdot (y_n - (1-\alpha))$, or explicity:

$$\theta_{n+1} = \begin{cases} \theta_n - \frac{c\alpha}{n} & \text{if } y_n = 1 \\ \theta_n + \frac{c(1-\alpha)}{n} & \text{if } y_n = 0 \end{cases} \tag{7}$$

The procedure is shown to be fully asymptotic efficient if $c = 1/g'(u_\theta)$. However, as neither g nor u_θ are known in advance, c is estimated adaptively, using the current estimate θ_n in place of u_θ, and assuming a parametric form for g [36].

The modified Robbins-Monro procedure. As mentioned above, if the optimal step size constant is known, this procedure is fully asymptotic efficient. However it was empirically shown to work poorly for extreme quantiles. Joseph [39] suggested a modification of this procedure, which is tuned to obtain optimal convergence speed. It uses the following update form:

$$\theta_{n+1} = \theta_n - a_n(y_n - C_n). \tag{8}$$

Joseph allows the use of a different target value, C_n, in each iteration, instead of the required constant, C. The step sizes a_n and target values C_n are derived explicitly in [39] to be optimal under a Bayesian analysis framework. As in [36],

the optimal step size also uses $g'(u_\theta)$, which is unknown, and a suitable approximation scheme is used. The modified Robbins-Monro procedure achieves significantly faster convergence rates in the case of the estimation of extreme quantiles.

3.4 Using the Modified Robbins-Monro Procedure to Obtain CIs for Heritability

We now describe how to rapidly construct CIs for heritability. As described above, the first step is to find s, t, s^* and t^*. To find s, we employ the modified Robbins-Monro procedure [39], where the parameter of interest is $\theta := h^2$, the function is $g(\theta) := \Pr_{h^2=\theta}(\hat{h}^2 = 0)$ and the inverse value we wish to find corresponds to $C = \alpha/2$. We note that we chose g here to be nonincreasing for the sake of clarity of presentation; to conform with the Robbins-Monro formulation, we would need to redefine $g \to 1 - g$ and $C \to 1 - C$. At a single iteration of the modified Robbins-Monro procedure, we have an estimate h_n^2 for s, and we need to sample from a distribution whose mean is $\Pr_{h_n^2}(\hat{h}^2 = 0)$. To achieve that, we draw a sample from the distribution corresponding to h_n^2, $\mathcal{N}\left(\mathbf{0}_n, \mathbf{V}_{h_n^2}\right)$, and check if the maximum likelihood estimate for it is 0 (or above). This procedure can be done quickly in $O(n)$, as we now describe, circumventing the need to perform a full likelihood maximization for the sample.

As detailed above, we make repeated use of the following procedure: (1) Draw a random sample $\mathbf{y}$ from the distribution correpsonding to a given heritability value h^2, $\mathcal{N}(\mathbf{0}_n, \mathbf{V}_{h^2})$; (2) Decide whether its heritability estimate, $\hat{h}^2(\mathbf{y})$, is larger than a given value, H^2. In [20], it is shown that when $\mathbf{X} = \mathbf{1}_n$, these two steps may equivalently be performed by drawing a vector $\mathbf{u}$ of i.i.d, standard normal variables $\mathbf{u} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$, and checking if

$$\sum_{i=1}^{n} \xi_i^{h^2,H^2} u_i^2 > 0 \ , \tag{9}$$

where

$$\xi_i^{h^2,H^2} = \frac{h^2(d_i - 1) + 1}{H^2(d_i - 1) + 1} \left(\frac{d_i - 1}{H^2(d_i - 1) + 1} - \frac{1}{n-1} \sum_{j=1}^{n-1} \frac{d_j - 1}{H^2(d_j - 1) + 1} \right) \tag{10}$$

for $i = 1, \ldots, n-1$, and $\xi_n^{h^2,H^2} = 0$, with d_i being the eigenvalues of $\mathbf{K}$. The sign of the expression in Equation (9) is equal to the sign of $\frac{\partial \ell_{REML}}{\partial h^2}(H^2)$, the derivative of ℓ_{REML} at the point H^2. Therefore, assuming the restricted likelihood function is well behaved, a positive derivative indicates that the REML heritability estimate is larger than H^2. Similar expressions are defined for a general $\mathbf{X}$ in [20]. Once the eigendecomposition of $\mathbf{K}$ is obtained, this procedure may be performed in a time complexity linear in n.

Similarly, for finding s^*, we define the function $g(\theta) := \Pr_{h^2=0}(\hat{h}^2 \leq \theta)$, for which we want to find the inverse of $C = 1 - \alpha$. The procedures for finding t and t^* are similar.

The second step involves calculating the quantities $c^{-1}_{\alpha/2}(H^2), c^{-1}_{\alpha}(H^2), c^{-1}_{1-\alpha}(H^2)$ and $c^{-1}_{1-\alpha/2}(H^2)$ as required. This can again be done by the modified Robbins-Monro procedure, by setting $\theta := h^2$, $g(\theta) := \Pr_{h^2}(\hat{h}^2 \leq H^2)$, and $C = \alpha/2, \alpha, 1-\alpha/2$ or $1-\alpha$. To sample from a distribution whose mean is $\Pr_{h^2_n}(\hat{h}^2 \leq H^2)$, we draw a sample from the distribution corresponding to h^2_n, and check if the maximum likelihood estimate for it is above H^2. Again, this procedure can be done quickly in $O(n)$. Once these quantities have been calculated, the CI can be calculated as detailed in Equations (3) and (4).

In practice, we used the following choices in the modified Robbins-Monro procedure: (i) We used $T = 1000$ iterations; (ii) we set the prior standard deviation to $\tau = 0.4$, used to derive a_n and C_n via the Bayesian analysis (see [39]); (iii) we used the midpoint between the estimate and relevant boundary (0 or 1, depending on the quantile required) as a starting point; (iv) we adaptively changed the step size constant, following the suggestion of Garthwaite and Buckland, by approximating the derivative with an expression proportional to the distance from $\hat{\theta}$:

$$g'(u_\theta) \approx k(h^2_n - H^2), \quad k = \frac{2}{z_\beta \cdot (2\pi)^{-1/2} \cdot e^{-z^2_\beta/2}} \tag{11}$$

where z is the quantile function of the normal distribution, and β is the required quantile.

3.5 The NFBC Dataset

We analyzed 5,236 individuals from the Northern Finland Birth Cohort (NFBC) dataset, which consists of genotypes at 331,476 genotyped SNPs and 10 phenotypes [37]. From each pair of individuals with relatedness of more than 0.025, one was reserved, resulting in 2,520 individuals.

3.6 The WTCCC2 Dataset

We analyzed the Wellcome Trust Case Control Consortium 2 dataset [38]. In the multiple sclerosis (MS) and ulcerative colitis (UC) datasets, we used the same data processing described in [44] to ensure consistency. Briefly, UK controls and cases from both UK and non-UK were used. SNPs were removed with $>$ 0.5% missing data, $p < 0.01$ for allele frequency difference between two control groups, $p < 0.05$ for deviation from Hardy-Weinberg equilibrium, $p < 0.05$ for differential missingness between cases and controls, or minor allele frequency $< 1\%$. In all analyses, SNPs within 5M base pairs of the human leukocyte antigen (HLA) region were excluded, because they have large effect sizes and highly unusual linkage disequilibrium patterns, which can bias or exaggerate the results. Finally, from each pair of individuals with relatedness of more than 0.025, one was reserved, resulting in 13,950 individuals.

4 Discussion

We have presented FIESTA, an efficient method for constructing accurate CIs using stochastic approximation. We have shown that FIESTA is very fast, while achieving exact coverage due to the fact that it does not rely on any assumptions of the distribution of the estimator. FIESTA is also faster than the analytical approximation used by SEs. Due to its speed, FIESTA can be easily used for datasets with tens or hundreds of thousands of individuals.

FIESTA requires the eigendecomposition of the kinship matrix, whose computational complexity is cubic in the number of individuals. While this is often a preliminary step in heritability estimation, it may be computationally prohibitive for larger datasets. Recent methods for heritability estimation (see [45]) utilize conjugate gradient methods to avoid cubic steps altogether. One direction of extension for FIESTA is devising a procedure to calculate the derivative of the restricted likelihood function using conjugate gradient methods, which are quadratic, but do not require the eigendecomposition.

We note that the confidence intervals constructed by FIESTA are estimated under a set of assumptions, particularly that the data is generated from the linear mixed model as described in the Methods. Deviations from these assumptions could result in inaccurate confidence intervals. Specifically, we observed that when the genotype matrix is of low rank (e.g., in the case where duplicates are introduced), then the confidence intervals calculated by FIESTA may be inaccurate. We therefore recommend removing duplicates and closely related individuals from the data prior to the application of FIESTA.

A common extension of the LMM is that of multiple variance components, where the genome is divided into distinct partitions (e.g., according to functional annotations, or by chromosomes), and the relative genetic contribution of each partition is estimated instead. Another extension is that of multiple traits, where several phenotypes are estimated concurrently, allowing dependencies between them. In principle, the methodology behind FIESTA can be applied to the multiparametric case as well. However, there are several computational and conceptual hurdles that make this application highly nontrivial. First, a major difficulty rises from the fact that it is no longer necessarily possible to jointly diagonalize several kinship matrices. Thus, the computation of the derivatives of the logarithm of the restricted likelihood functions can no longer utilize the eigendecomposition. Second, the inversion of acceptance regions of multiple parameters results in confidence regions of more than one dimension. While these have the required coverage probability, their shape may be difficult to report or to interpret easily (e.g. an ellipsoid). For example, hyper-rectangular confidence regions are often desirable [46], as the marginal CI of each parameter has the same coverage probability as the confidence region. Therefore, multiparametric extensions remain a future direction of research.

Acknowledgements. The authors would like to thank David Steinberg. R.S. is supported by the Colton Family Foundation. This study was supported in

part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel Aviv University to R.S. The Northern Finland Birth Cohort data were obtained from dbGaP: phs000276.v2.p1. This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from `www.wtccc.org.uk`. Funding for the project was provided by the Wellcome Trust under award 076113.

Appendix The supplementary material, including additional figures, are located at `https://github.com/cozygene/albi`.

References

1. Fisher, R.A.: The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh **52** (1918) 399–433
2. Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D.I., Cornes, B.K., Davis, C., Dunkel, L., De Lange, M., Harris, J.R., Hjelmborg, J.V., et al.: Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Research **6**(05) (2003) 399–408
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**(4) (2006) 571–580
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**(5) (2008) 1590
5. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., Parkinson, H.: The NHGRI GWAS catalog, a curated resource of snp-trait associations. Nucleic Acids Research **42**(Database issue) (1 2014) D1001–6
6. Visscher, P.M., Hill, W.G., Wray, N.R.: Heritability in the genomics era – concepts and misconceptions. Nature Reviews Genetics **9**(4) (2008) 255–266
7. Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., Eskin, E.: Efficient control of population structure in model organism association mapping. Genetics **178**(3) (3 2008) 1709–23
8. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y.Y., Freimer, N.B., Sabatti, C., Eskin, E.: Variance component model to account for sample structure in genome-wide association studies. Nature Genetics **42**(4) (4 2010) 348–54
9. Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., Heckerman, D.: Fast linear mixed models for genome-wide association studies. Nature Methods **8**(10) (2011) 833–5
10. Zhou, X., Stephens, M.: Genome-wide efficient mixed-model analysis for association studies. Nature Genetics **44**(7) (2012) 821–4
11. Vattikuti, S., Guo, J., Chow, C.C.: Heritability and genetic correlations explained by common snps for metabolic syndrome traits. PLoS Genetics **8**(3) (2012) e1002637
12. Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y.H., Abdellaoui, A., Batista, S., Butler, C., Chen, G., Chen, T.H., D'Ambrosio, D., Gallins, P., Ha, M.J., Hottenga, J.J., Huang, S., Kattenberg, M., Kochar, J., Middeldorp, C.M., Qu, A., Shabalin, A., Tischfield,

J., Todd, L., Tzeng, J.Y., van Grootheest, G., Vink, J.M., Wang, Q., Wang, W., Wang, W., Willemsen, G., Smit, J.H., de Geus, E.J., Yin, Z., Penninx, B.W.J.H., Boomsma, D.I.: Heritability and genomics of gene expression in peripheral blood. Nature Genetics **46**(5) (05 2014) 430–437
13. Kruijer, W., Boer, M.P., Malosetti, M., Flood, P.J., Engel, B., Kooke, R., Keurentjes, J.J., van Eeuwijk, F.A.: Marker-based estimation of heritability in immortal populations. Genetics **199**(2) (2015) 379–398
14. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M.: Common SNPs explain a large proportion of the heritability for human height. Nature Genetics **42**(7) (7 2010) 565–9
15. 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**(1) (2011) 76–82
16. Lohr, S.L., Divan, M.: Comparison of confidence intervals for variance components with unbalanced data. Journal of Statistical Computation and Simulation **58**(1) (1997) 83–97
17. Burch, B.D.: Comparing pivotal and REML-based confidence intervals for heritability. Journal of Agricultural, Biological, and Environmental Statistics **12**(4) (2007) 470–484
18. Burch, B.D.: Assessing the performance of normal-based and REML-based confidence intervals for the intraclass correlation coefficient. Computational Statistics & Data Analysis **55**(2) (2011) 1018–1028
19. Kraemer, K.: Confidence intervals for variance components and functions of variance components in the random effects model under non-normality. (2012)
20. Schweiger, R., Kaufman, S., Laaksonen, R., Kleber, M.E., März, W., Eskin, E., Rosset, S., Halperin, E.: Fast and accurate construction of confidence intervals for heritability. The American Journal of Human Genetics **98**(6) (2016) 1181–1192
21. Chernoff, H.: On the distribution of the likelihood ratio. The Annals of Mathematical Statistics (1954) 573–578
22. Moran, P.A.: Maximum-likelihood estimation in non-standard conditions. In: Mathematical Proceedings of the Cambridge Philosophical Society. Volume 70., Cambridge Univ Press (1971) 441–450
23. 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**(398) (1987) 605–610
24. Stern, S., Welsh, A.: Likelihood inference for small variance components. The Canadian Journal of Statistics **28**(3) (2000) 517–532
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**(1) (2015) 223–232
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**(3) (2013) 129–140
27. Wolfinger, R.D., Kass, R.E.: Nonconjugate Bayesian analysis of variance component models. Biometrics **56**(3) (2000) 768–774
28. Chung, Y., Rabe-hesketh, S., Gelman, A., Dorie, V., Liu, J.: Avoiding boundary estimates in linear mixed models through weakly informative priors. Berkeley Preprints (2011) 1–30
29. Harville, D.A., Fenech, A.P.: Confidence intervals for a variance ratio, or for heritability, in an unbalanced mixed linear model. Biometrics (1985) 137–152

30. Burch, B.D., Iyer, H.K.: Exact confidence intervals for a variance ratio (or heritability) in a mixed linear model. Biometrics (1997) 1318–1333
31. Furlotte, N.A., Heckerman, D., Lippert, C.: Quantifying the uncertainty in heritability. Journal of Human Genetics **59**(5) (2014) 269–275
32. Carpenter, J., Bithell, J.: Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. Statistics in Medicine **19**(9) (2000) 1141–1164
33. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al.: Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med **12**(3) (2015) e1001779
34. Kushner, H., Yin, G.G.: Stochastic approximation and recursive algorithms and applications. Volume 35. Springer Science & Business Media (2003)
35. Robbins, H., Monro, S.: A stochastic approximation method. The annals of mathematical statistics (1951) 400–407
36. Garthwaite, P.H., Buckland, S.T.: Generating monte carlo confidence intervals by the robbins-monro process. Applied Statistics (1992) 159–171
37. Sabatti, C., Service, S.K., Hartikainen, A.L.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M., Sovio, U., Ruokonen, A., Laitinen, J., Jakkula, E., Coin, L., Hoggart, C., Collins, A., Turunen, H., Gabriel, S., Elliot, P., McCarthy, M.I., Daly, M.J., Järvelin, M.R.R., Freimer, N.B., Peltonen, L.: Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nature Genetics **41**(1) (1 2009) 35–46
38. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E., et al.: Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature **476**(7359) (2011) 214
39. Joseph, V.R.: Efficient robbins–monro procedure for binary data. Biometrika **91**(2) (2004) 461–470
40. Furlotte, N.A., Eskin, E.: Efficient multiple trait association and estimation of genetic correlation using the matrix-variate linear mixed-model. Genetics **200**(1) (2015) 59–68
41. Searle, S.R., Casella, G., McCulloch, C.E.: Variance components. Volume 391. John Wiley & Sons (2009)
42. Visscher, P.M., Hill, W.G., Wray, N.R.: Heritability in the genomics eraconcepts and misconceptions. Nature Reviews Genetics **9**(4) (2008) 255–266
43. Patterson, H.D., Thompson, R.: Recovery of inter-block information when block sizes are unequal. Biometrika **58**(3) (1971) 545–554
44. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., Price, A.L.: Advantages and pitfalls in the application of mixed-model association methods. Nature genetics **46**(2) (2014) 100–106
45. Loh, P.R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., of the Psychiatric Genomics Consortium, S.W.G., de Candia, T.R., Lee, S.H., Wray, N.R., Kendler, K.S., O'Donovan, M.C., Neale, B.M., Patterson, N., Price, A.L.: Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nature Genetics **47**(12) (12 2015) 1385–1392
46. Sidak, Z.: Rectangular confidence regions for the means of multivariate normal sistributions. Journal of the American Statistical Association **62**(318) (1967) 626–633

47. Wasserman, L.: All of statistics: a concise course in statistical inference. Springer Science & Business Media (2013)
48. Gilmour, A.R., Thompson, R., Cullis, B.R.: Average information reml: an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics (1995) 1440–1450

5 Appendix

5.1 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 [47]. The observed information matrix $\mathcal{J}(\boldsymbol{\theta})$ of parameters $\boldsymbol{\theta}$ is the negative of the Hessian of the log-likelihood of the data $\mathbf{y}$. Namely, $\mathcal{J}(\boldsymbol{\theta})_{i,j} = -\frac{\partial}{\partial\theta_i\theta_j}\ell(\boldsymbol{\theta};\mathbf{y})$. The Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$ is the expectation of the observed information matrix. Namely, $\mathcal{I}(\boldsymbol{\theta})_{i,j} = \mathrm{E}\left[-\frac{\partial}{\partial\theta_i\theta_j}\ell(\boldsymbol{\theta};\mathbf{y})\right]$. Asymptotically, under certain regularity conditions, $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta})^{-1})$. According to the delta method, the asymptotic distribution of a function $f(\boldsymbol{\theta})$ satisfies $\sqrt{n}(f(\widehat{\boldsymbol{\theta}}) - f(\boldsymbol{\theta})) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \nabla f(\boldsymbol{\theta})^T \mathcal{I}(\boldsymbol{\theta})^{-1} \nabla f(\boldsymbol{\theta}))$.

GCTA uses the Average Information [48] (AI) matrix $\mathcal{A}$ to calculate the variance of σ_g^2 and σ_e^2, where $\mathcal{A} = \frac{1}{2}(\mathcal{I} + \mathcal{J})$. For the REML method, this is the matrix:

$$\mathcal{A} = \frac{1}{2} \cdot \begin{pmatrix} \mathbf{y^T QKQKQy} & \mathbf{y^T QKQQy} \\ \mathbf{y^T QQKQy} & \mathbf{y^T QQQy} \end{pmatrix}, \tag{12}$$

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

$$\mathrm{Var}(\hat{h}^2) = (\hat{\sigma}_g^2 + \hat{\sigma}_e^2)^{-4}\left(\hat{\sigma}_e^2 - \hat{\sigma}_g^2\right)\mathcal{A}^{-1}|_{\sigma_g^2=\hat{\sigma}_g^2,\sigma_e^2=\hat{\sigma}_e^2}\begin{pmatrix}\hat{\sigma}_e^2 \\ -\hat{\sigma}_g^2\end{pmatrix}. \tag{13}$$

Given the eigendecomposition of $\mathbf{K}$, $\boldsymbol{\Sigma}^{-1}$ (and thus $\mathbf{Q}$) can be calculated in $O(n)$ (where n is the number of individuals), avoiding an expensive matrix inversion. Several other computational improvements may be carried out, depending on software implementation. However, we note that $O(n^2)$ matrix-by-vector multiplications cannot be avoided.

5.2 Confidence intervals for heritability

Our approach is based on the duality between hypothesis testing and confidence intervals. As the distribution of $\hat{h}^2$ depends solely on h^2, we may assume without loss of generality that $\sigma_p^2 = 1$ and $\boldsymbol{\beta} = \mathbf{0}_p$. 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 vector is drawn from $\mathcal{N}(\mathbf{0}_n, \mathbf{V}_{h^2})$. The probability of the event A_{h^2} under $\mathcal{N}(\mathbf{0}_n, \mathbf{V}_{h^2})$ 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 for an estimate $\hat{h}^2 = H^2$, C_{H^2}, comprises of the set of parameter values for which

$\hat{h}^2$ does not imply the rejection of the null hypothesis that the true heritability value is h^2:

$$C_{H^2} = \left\{ h^2 \,\middle|\, H^2 \in A_{h^2} \right\} \ . \tag{14}$$

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 [20], 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; i.e. $\Pr_{h^2}(\hat{h}^2 \leq c_\beta(h^2)) = \beta$. A natural choice for A_{h^2} would be taking the interval obtained by removing a $\alpha/2$-tail from both sides of the distribution of $\hat{h}^2$ given h^2, i.e., choosing the two-sided $A_{h^2} = [c_{\alpha/2}(h^2), c_{1-\alpha/2}(h^2)]$. If this were always possible, a succinct way of describing the $1-\alpha$ CI, $C_{H^2} = [l_{H^2}, h_{H^2}]$, would be using the fact that its endpoints are exactly those following

$$c_{1-\alpha/2}(l_{H^2}) = H^2 \Rightarrow l_{H^2} = c^{-1}_{1-\alpha/2}(H^2) \tag{15}$$

$$c_{\alpha/2}(h_{H^2}) = H^2 \Rightarrow u_{H^2} = c^{-1}_{\alpha/2}(H^2). \tag{16}$$

as described in Figure 3.

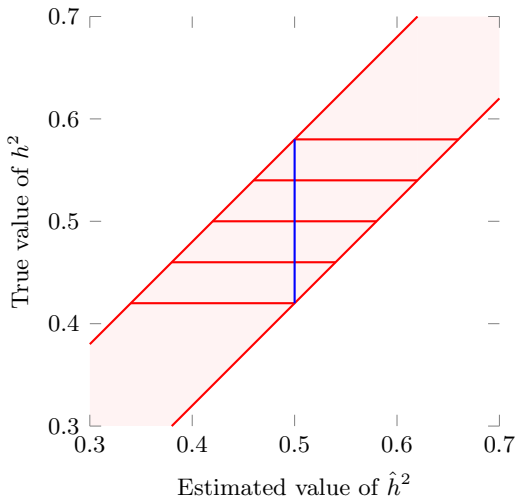

Fig. 3: **An illustration of acceptance regions and CIs.** The diagonal lines are the $\alpha/2$ and $1-\alpha/2$ quantile functions, shown for values in the mid-range of heritability values. Several example acceptance regions are denoted as horizontal lines, in parameter regions where simple two-sided acceptance regions can be defined. The CI for $\hat{h}^2 = 0.5$ is shown as a vertical line.

However, since the distribution is of a mixed type with discontinuity points, it may be the case that the probability of the interval $[c_{\alpha/2}(h^2), c_{1-\alpha/2}(h^2)]$ might be greater than $(1-\alpha/2)-\alpha/2 = 1-\alpha$. For example, if $\Pr_{h^2}(\hat{h}^2 = 0) > \alpha/2$, then $c_{\alpha/2} = 0$, and $\Pr_{h^2}(\hat{h}^2 \in [0, c_{\alpha/2})) > \alpha/2$. In this case, we then instead choose to take the one-sided interval $A_{h^2} = [0, c_{1-\alpha}(h^2)]$. Similarly, if $\Pr_{h^2}(\hat{h}^2 = 1) > \alpha/2$, then $c_{1-\alpha} = 1$, and $\Pr_{h^2}(\hat{h}^2 \in (c_{1-\alpha/2}, 1]) > \alpha/2$. In this case, we similarly choose the one-sided interval $[c_\alpha(h^2), 1]$ instead. We are therefore interested in the maximal value s for which $\Pr_s(\hat{h}^2 = 0) \geq \alpha/2$, and the minimal value t for which $\Pr_{h^2}(\hat{h}^2 = 1) \geq \alpha/2$, because in the range of values $h^2 \in [s, t]$, it holds that $\Pr_{h^2}(\hat{h}^2 \in [c_{\alpha/2}(h^2), c_{1-\alpha/2}(h^2)]) = 1 - \alpha$. Equivalently, assuming $\Pr_{h^2}(\hat{h}^2 = 0)$ (resp., $\Pr_{h^2}(\hat{h}^2 = 1)$) is decreasing (resp., increasing) in h^2, we may simply define s and t to be the values for which $\Pr_{h^2=s}(\hat{h}^2 = 0) = \alpha/2$ and $\Pr_{h^2=t}(\hat{h}^2 = 1) = \alpha/2$.

The following assumes s and t exist, and that $s < t$; for the general case, see [20]. We divide our construction into distinct cases, by setting

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

The three region types are illustrated by Figure 4. Inverting the acceptance regions, we get the following definition for $C_{H^2} = [l_{H^2}, h_{H^2}]$. For the lower endpoint, we have

$$l_{H^2} = \begin{cases} 0 & \text{if } H^2 \in [0, c_{1-\alpha}(0)) \\ c^{-1}_{1-\alpha}(H^2) & \text{if } H^2 \in [c_{1-\alpha}(0), c_{1-\alpha}(s)) \\ s & \text{if } H^2 \in [c_{1-\alpha}(s), c_{1-\alpha/2}(s)) \\ c^{-1}_{1-\alpha/2}(H^2) & \text{if } H^2 \in [c_{1-\alpha/2}(s), 1] \end{cases}$$

For the higher endpoint, we have

$$u_{H^2} = \begin{cases} c^{-1}_{\alpha/2}(H^2) & \text{if } H^2 \in [0, c_{\alpha/2}(t)) \\ t & \text{if } H^2 \in [c_{\alpha/2}(t), c_\alpha(t)) \\ c^{-1}_{\alpha}(H^2) & \text{if } H^2 \in [c_\alpha(t), c_\alpha(1)) \\ 1 & \text{if } H^2 \in [c_\alpha(1), 1] \end{cases}$$

These conditions, phrased in terms of the quantile functions c_β, e.g., $H^2 \leq c_\alpha(t)$, can be equivalently written in terms of the value of inverse quantile functions of the estimate H^2, e.g. $c^{-1}_\alpha(H^2) \leq t$. In addition, let $s^* = c_{1-\alpha}(0), t^* = c_\alpha(1)$. Explicitly,

$$l_{H^2} = \begin{cases} 0 & \text{if } H^2 \leq s^* \\ c^{-1}_{1-\alpha}(H^2) & \text{if } c^{-1}_{1-\alpha}(H^2) < s \\ s & \text{if } s \in [c^{-1}_{1-\alpha/2}(H^2), c^{-1}_{1-\alpha}(H^2)] \\ c^{-1}_{1-\alpha/2}(H^2) & \text{if } s < c^{-1}_{1-\alpha/2}(H^2) \end{cases} \tag{17}$$

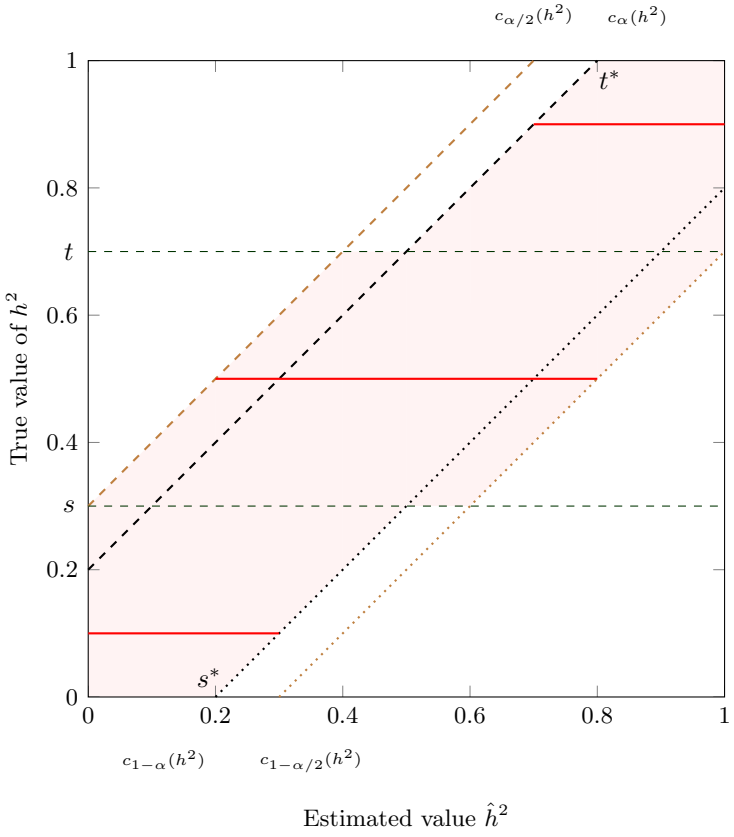

Fig. 4: **An illustration of the three acceptance regions types.** The diagonal lines, from left to right, indicate the quantile functions for $\alpha/2, \alpha, 1-\alpha$ and $1-\alpha/2$. The three region types are indicated as horizontal lines. The points s and t, where region types used are changed, are indicated as horizontal dashed lines. See Methods for a full description.

and

$$u_{H^2} = \begin{cases} c_{1-\alpha/2}^{-1}(H^2) & \text{if } c_{\alpha/2}^{-1}(H^2) < t \\ t & \text{if } t \in [c_{\alpha}^{-1}(H^2), c_{\alpha/2}^{-1}(H^2)] \\ c_{\alpha}^{-1}(H^2) & \text{if } t < c_{\alpha}^{-1}(H^2) \\ 1 & \text{if } t^* \leq H^2 \end{cases} \tag{18}$$

It follows from the discussion above, that in order to construct a CI for an heritability estimate H^2, we need to first find s, t as above, $s^* = c_{1-\alpha}(0)$ and $t^* = c_{\alpha}(1)$, and then we need only calculate $c_{\beta}^{-1}(H^2)$ for $\beta = \alpha/2, \alpha, 1-\alpha$ and $1-\alpha/2$. Therefore, the entire construction relies on inverting certain quantile functions.

5.3 Accuracy of ALBI CIs

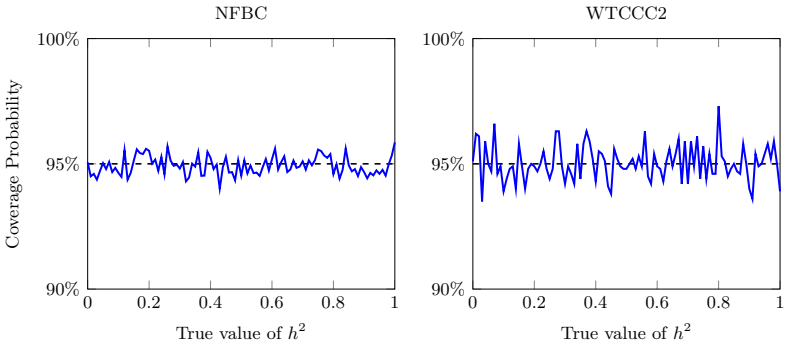

Fig. 5: **Accuracy of CIs for the NFBC and WTCCC2 datasets.** The coverage probabilities of the ALBI CIs. The coverage probabilities are shown for CIs designed to have coverage probabilities of 95%.