ON NEGATIVE HERITABILITY AND NEGATIVE ESTIMATES OF HERITABILITY

DAVID STEINSALTZ, ANDY DAHL, AND KENNETH W. WACHTER

Abstract. We consider the problem of interpreting negative maximum likelihood estimates of heritability that sometimes arise from popular statistical models of additive genetic variation. These may result from random noise acting on estimates of genuinely positive heritability, but we argue that they may also arise from misspecification of the standard additive mechanism that is supposed to justify the statistical procedure. Researchers should be open to the possibility that negative heritability estimates could reflect a real physical feature of the biological process from which the data were sampled.

1. Introduction: The meaning of heritability

1.1. Operational definitions of heritability.

As Albert Jacquard [9] pointed out decades ago, *narrow-sense heritability* — commonly denoted h^2 — has conventionally two distinct meanings:

1. The proportion of total variance attributable to additive genetic effects;
2. The slope of the linear regression of children's phenotypes on the mean parental phenotypes.

Both meanings appear in the earliest works to give a quantitative operational definition to *heritability*, in particular [12]. (For more about the history of the notion of heritability see [2].)

The correspondence between these two meanings depends on an additive model, where genetic and non-genetic effects are independent and sum together to produce the phenotype. When we have general genetic relatedness (rather than parental relations with fixed 50% relatedness) heritability is analogous to a regression coefficient relating phenotypic similarity to genotypic similarity.

We are particularly concerned here with the interpretation of negative estimates of heritability. The appearance of negative estimates for a parameter of crucial scientific interest that is *prima facie* positive is unusual, as has often been noted. Negative estimates of the heritability parameter are often dismissed as a mathematical abstraction, values in a range that arises purely formally and that may only be reported for formal purposes, as part of an ensemble of estimates that collectively are unbiased. Several recent studies [19, 3, 5, 21] have reported individual negative heritability estimates in this way, including

DS supported by Grant ES/N011856/1 from the UK Economic and Social Research Council. AD supported by Grant 1U01HG009080-01 from the U.S. National Institute of Health. KWW supported by Grant 5P30AG012839 from the U.S. National Institute on Aging.

them in averages that themselves came out positive. In [10] a point estimate of -0.109 is obtained for heritability of horn length in Soay sheep. It is immediately dismissed with the statement that "it is impossible to have negative heritability" and the inference is drawn that the true heritability must actually be a small positive number toward the upper end of the confidence interval.

We wish to argue that negative heritability estimates need to be taken more seriously. The confusion, we contend, comes from the overlap between the statistical models that operationalise the two different interpretations of heritability described above. The argument for rejecting negative estimates appears compelling just so long as the focus is only on the random-effects probability model 1. Variance is nonnegative, hence the ratio of two variances cannot be negative. The denominator represents total variance and the numerator represents one component of variance, implying a ratio in $[0, 1]$ if the two components are independent (as the model presumes).

While "variance attributable to additive genetic effects" is a basic element of the genetic model, it has no place in the statistical algorithms such as GREML derived from this model that is widely used to estimate heritability from experimental data. The GREML algorithm is actually (as we will explain in section 2.1) the realization of a multivariate normal model that is naturally constrained to have the parameter $h^2 \geq -1/(\max\{s_i^2\} - 1)$, where $(s_i)_{i=1}^n$ are the singular values of the genotype matrix. If the phenotypes were derived from summing independent additive genetic effects then the true h^2 must indeed be nonnegative, but that must be recognized as an additional assumption that must be scientifically warranted, as it is not compelled on any formal grounds.

1.2. **The meaning of negative heritability.** Once we have accepted the GREML multivariate normal framework–which we will define precisely–we must admit the possibility that the joint distribution of phenotypes and genotypes in a given dataset may be best described by an h^2 value that is negative. The question this raises is, can such a negative heritability estimate be biologically sensible? As described in Section 2, the parameter for heritability may be identified, in a precise way, with a correlation between genotype similarity and phenotype similarity. The model invites us to select an estimate of h^2 that will best match the genetic covariance between individuals to the similarity in their traits. Even if we *want* heritability to be interpreted in the first sense, as a partition of variance, this will not, in general, be correct. All we have access to from the data is an estimate of something like heritability in sense 2. High heritability means that individuals with similar genotype are likely to have similar trait values. Zero heritability means that genotypes tell us nothing about similarities in trait values. Negative heritability, then, could be perfectly sensible as a description of the data: It means that individuals with similar genotypes are likely to have more divergent trait values than those with highly disparate genotypes.

Saying that a given set of data might be best described by a negative heritability estimate goes only part of the way toward answering the question of the biological plausibility of the concept. Suppose you were estimating the weight of water droplets by successively adding them to a small container, and estimating the slope from the sequence of weights. If the scale is sufficiently imprecise it is hardly unlikely that we could estimate a negative slope,

yet common sense tells us that negative estimates should be dismissed as unrealistic, and truncated at 0. Statistical theory tells us that this system should produce slope estimates that may be positive or negative, but that the probability of a negative estimate goes to 0 as the number of measurements goes to infinity, and the estimate converges to its true (positive) value. The essential question is, is there a plausible mechanism that could produce genuine negative heritability, so that as the amount of data generated by the model goes to infinity, the estimate converges to a negative quantity.

The term "negative heritability" appeared for the first time, so far as we are aware, in a paper [8] by J. B. S. Haldane, written around 1960, but first published posthumously in 1996. Haldane described how the maternal-effect trait of neonatal jaundice could be said to display negative heritability: Because the disease results from maternal antibodies against a fetal antigen, it will not arise in a fetus whose mother herself experienced neonatal jaundice.[1] Haldane then calculates a negative heritability from a model that is specialized to the peculiar structure of neonatal jaundice.

We will suggest one such mechanism in Section 4. As with Haldane's model (which may be understood as a special case), this mechanism has implications which may be implausible or even obviously false in a given experimental setting. It involves interactions between individuals that are not primarily genetic, and so may be dismissed as irrelevant to the study of genetic heritability. The point we want to suggest, though, is that as an abstract physical mechanism that could be producing our data it is as mathematically plausible as the linear random-effects model that undergirds GREML. This is only one example of such a mechanism, and the conclusion we wish to suggest is that negative heritability must be acknowledged as a genuine phenomenon for genotype-phenotype data, even if it may be reasonably excluded by the context of some particular studies. Thinking about what sorts of biological settings could yield negative heritability can also prove an effective guide to understanding when negative heritability estimates may be reliably truncated or ignored.

This is very much like the advice on "interpretation of negative components of variance" propounded in a very different context by the statistician J. A. Nelder [13] in 1954. Nelder considered the problem of ANOVA testing on split-plot experiments, where error for main plots was found to be smaller than the error for subplots, producing a negative estimate for the residual subplot error. As we have done here, Nelder showed how the apparently negative "variance component" could arise either from sampling error from a positive variance component, or from a misspecification of the model, where correlations between measurements have been neglected. "In any particular situation," Nelder concludes, "it is the statistician's responsibility to decide which model is more appropriate."

2. The GREML model as linear regression

2.1. **The random-effects model.** For the remainder of this paper we follow [14] in using the letter ψ to represent heritability, to avoid the confusing implication built in to the nomenclature h^2 that this parameter cannot be negative.

[1] We thank Jonathan Marchini for pointing out this reference to us.

Underlying GREML, but also alternative approaches to heritability estimation such as LD-score and Haseman-Elston regression, is a basic random-effects model. Following the notation of [14], our basic object is a data set consisting of an $n \times p$ matrix Z, taken to represent the genotypes of n individuals, measured at p different loci. There is a vector $\mathbf{y}$, representing a scalar observation for each of the n individuals. The underlying observations are counts of alleles taking the values $0, 1$, or 2, but the genotype matrix is centered to have mean zero in each column and normalized to have mean square over the whole matrix equal to 1 (often, columns are further standardized to variance 1). The model posits the existence of a random vector $\mathbf{u} \in \mathbb{R}^p$ of genetic influences from the individual SNPs such that

$$\mathbf{y} = Z\mathbf{u} + \varepsilon. \tag{1}$$

The vectors $\mathbf{u}$ and ε are assumed to be independent and to have zero means and i.i.d. normal components. The variances are determined by two parameters, which are to be estimated: θ represents the precision (reciprocal variance) of the non-genetic noise and ψ represents the heritability, entering the model as the ratio of genetic variance to total variance. We will also use the notation $\phi = \psi/(1-\psi)$ in some places for concision.

The GREML model has been formulated as a random-effects model, but it is equivalent to a multivariate normal model corresponding to the covariance matrix

$$C^2 := \theta_0^{-1}\left((\psi/(1-\psi))\, A + I_n\right). \tag{2}$$

In this section we describe how the model may also be understood as a linear regression model. In their original paper [18], Yang and coauthors spell out an analogy between GCTA and a different form of linear regression. They regress squared trait differences between pairs of individuals on corresponding elements of the Genetic Relatedness Matrix, with $n(n-1)/2$ points and correlated errors (this is Haseman–Elston regression, which has recently become a popular heritability estimation method due to its speed and robustness to some degree of model misspecification [7, 4]). Instead, we draw an approximate comparison between GREML and regression with n points and independent errors.

Let $Z = U\mathrm{diag}(s_i)V^*$ be the singular-value decomposition of $Z/\sqrt{p}$, and rotate the observations to diagonalize the covariance matrix, obtaining

$$\mathbf{z} := U^*\mathbf{y}.$$

The elements of $\mathbf{z}$ are independent centered normal random variables, and $\mathbf{z}_i$ has variance $(1-\psi+\psi s_i^2)/(\theta(1-\psi))$. Define

$$w_i(\psi) := \frac{1-\psi}{1-\psi+\psi s_i^2}$$

and

$$v_i(\psi) := \frac{(1-\psi)z_i^2}{1-\psi+\psi s_i^2},$$

We also define $\tau_2(\psi) = \psi^{-2}\mathrm{Var}(w(\psi))$, and omit the dependence on ψ when helpful.

It was shown in [14] that the maximum-likelihood estimator of ψ may be written as the solution to the equation

$$0 = \mathrm{Cov}\left(\mathbf{w}(\hat{\psi}), \mathbf{v}(\hat{\psi})\right),$$

where Cov is to be understood as an operation on vectors, so $\mathrm{Cov}(\mathbf{x}, \mathbf{y}) := n^{-1}\sum(x_i - \bar{x})(y_i - \bar{y})$. Under the GREML model, the $\theta_0 v_i(\psi_0)$ are i.i.d. chi-squared random variables with one degree of freedom.

When $\psi(s_i^2 - 1)$ are uniformly small we may write this model as

$$\begin{aligned} \log z_i^2 &= -\log\hat{\theta} + \log\left(1 + \frac{\psi}{1-\psi}s_i^2\right) + \log v_i \\ &\approx -\left(\log\hat{\theta} + \psi + \log(1-\psi)\right)\Big) + \psi s_i^2 + \log v_i. \end{aligned} \tag{3}$$

If we compare this to a standard linear regression problem (with $\log z_i^2$ as the dependent variable, s_i^2 as the independent variable, and $\log v_i$ as the noise), we would expect to have variance of the slope estimate inversely proportional to the variance of the independent variables. That is, when the s_i^2 are tightly clustered, there will be large errors in the estimate of the slope, which is the parameter that gives information about ψ.

In addition, we would have to cope with the fact that the noise term is not normal, but log chi-square. This is highly left-skewed, with a very short tail on the positive side, and a long tail (asymptotically exponential) on the negative side.

2.2. **Simulations.** We plot in Figure 1 an example based on dimensions similar to those for the genotype matrix in the celebrated paper [18], but drawing singular values from the independent setting, not from the (unreported) empirical distribution of singular values underlying that study. For $n = 4000$ and $p = 100,000$, we show a scatterplot of the pairs (z_i^2, s_i^2) obtained from the singular-value decomposition of a genotype matrix. Note that the lines corresponding to disparate ψ estimates are very similar, and have little leverage relative to the huge scatter in the values of z_i^2.

In the independent setting, the known limiting measure for the singular values, the Marcenko–Pastur distribution (see section 5 of [14]), simplifies the task of exploring the GREML model through simulations. Instead of performing matrix multiplications and diagonalizations on a random $n \times p$ matrix, where n and p may be on the order of 10^5 or 10^6, we may instead start with the singular values. These are n equally spaced values from the singular value distribution, to which we contribute n normally distributed random variables z_i with mean zero and variance $1 + \frac{\psi_0 s_i^2}{1-\psi_0}$. We solve for $\hat{\psi}$ that makes the vector (z_i) and the vector (w_i) uncorrelated. That is, $\mathrm{Cov}(z_i(\psi), w_i(\psi))$ is a univariate function of ψ that is easy to compute, and will typically cross zero exactly once. (The exceptions are when the covariance is strictly negative for all ψ, meaning that the likelihood is increasing, so that the MLE is $\hat{\psi} = 1$; and when there are multiple solutions.)

This is essentially equivalent to simulating a random genotype matrix with i.i.d. entries and a random trait vector. (Furthermore, we could simulate a different model for the GRM simply by choosing a different distribution for the singular values.)

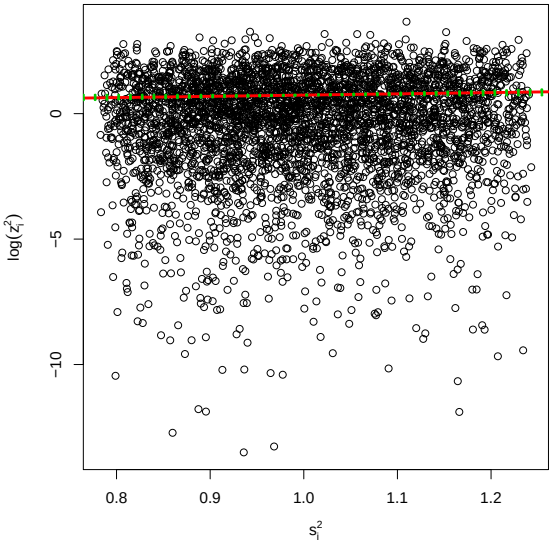

FIGURE 1. Linear-regression approximation of the estimation problem for ψ. The solid red line shows the correct line, corresponding to $\psi = 0.5$, while the dashed green line corresponds to $\psi = 0.25$.

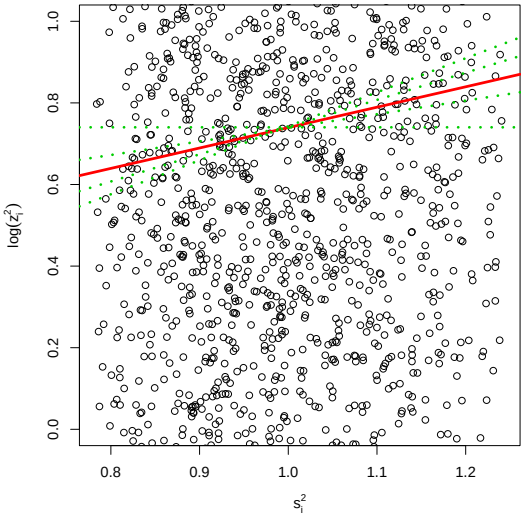

FIGURE 2. Blow-up of the linear-regression approximation, showing just the range $z_i^2 \in [0, 1]$. The dashed green lines correspond to $\psi = 0, 0.33, 0.67, 0.84$.

3. Bias from rejecting negative heritability estimates

The common practice of truncating the maximum likelihood calculation to non-negative values introduces bias that is well-known and may be serious for samples of moderate size, both when estimates are truncated at zero and when negatives are ignored. It is thus worth looking beyond the original motivation to the actual structure of the GREML model and considering what meaning negative parameter values might turn out to have.

The problem of estimating the probability of negative heritability estimates was studied fifty years ago by [6]. We add here a few comments about how the framework described in [14] may contribute to understanding the magnitude of the negative heritability estimate problem that arises from sampling noise in settings where the true heritability is understood to be nonnegative, hence where truncation at 0 (or rejection of negative estimates) is warranted and guarantees improved estimates in, say, MSE. We gain a rough idea of the impact of rejecting negative estimates from a normal approximation

$$\hat{\psi} - \psi_0 \approx \frac{\sqrt{2}}{\sqrt{n\,\tau_2}} \psi_0(1-\psi_0)X,$$

where X has standard normal distribution (see [14] for derivation). Rejecting estimates where $\hat{\psi} < 0$, we have the conditioning bias

$$\mathbb{E}[\hat{\psi}] - \psi_0 \approx \frac{(1-\psi_0)}{\sqrt{n\tau_2}} z_0 \frac{e^{-z_0^2/2}}{\sqrt{\pi}\Phi(z_0)}, \tag{4}$$

where $z_0 = \sqrt{n\tau_2}/\sqrt{2}(1-\psi_0)$ and Φ is the standard normal c.d.f. If we instead truncate the estimates — raising all negative $\hat{\psi}$ to 0 — we get the truncation bias

$$\mathbb{E}[\hat{\psi}] - \psi_0 \approx \frac{(1-\psi_0)}{\sqrt{n\tau_2}} z_0 \frac{e^{-z_0^2/2}}{\sqrt{\pi}} - \psi_0\big(1-\Phi(z_0)\big). \tag{5}$$

4. The phenotypic repulsion model

The notion that new species force their way into phenotypic gaps in the existing ecological community was termed by Darwin the "principle of divergence" and has been further developed by ecologists under the name "phenotypic repulsion" or "phylogenetic repulsion" [16]. Species living in close proximity — which are often closely related phylogenetically — coexist by separating from each other phenotypically. A similar kind of competitive exclusion has been proposed [15] on the individual level to explain observed pattern of developmental variation within human families. Social niche-formation within families has also been proposed by [1] — without an explicit mathematical model — as the basis for an evaluation of gene-environment interaction based on misclassified twin types. While we are not aware of mathematical models of this phenomenon, one could certainly imagine local competition for sunlight, combined with range-limited seed dispersion, yielding an effective phenotypic repulsion between related plants in a forest setting, or monozygotic twins who seek to distinguish themselves from their sibling.

We propose a model of phenotypic repulsion where individuals that are most closely related genetically strive to avoid each other phenotypically. We begin with a model like that described in Section 2.1, where individuals have phenotypes determined by normally distributed effect sizes acting on their individual genotypes. We introduce a penalty term to the probability, of the form

$$\exp\Big\{-\alpha\theta_0 \sum_{1\le i<j\le n} a_{ij} y_i y_j\Big\}$$

where $a_{ij} = \frac{1}{p}\sum_{k=1}^{p} Z_{ik}Z_{jk}$ is the (i,j) entry of the GRM, and $\alpha \le 1$ is a parameter measuring the extent to which genetically similar individuals are pushed to have differing phenotypes. Of course, this could be generalized to higher-dimensional phenotypes, with $y_i y_j$ replaced by an arbitrary inner product. The penalty term is inspired by the statistical mechanics models that have been applied to geographically-structured population dynamics, such as the Contact Process [11], used to model the spread of epidemics.

Combining this with (2) we see that the phenotypes will now be multivariate normal with mean 0 and covariance matrix

$$\theta_0^{-1}\Big[(\phi_0 A + I_n)^{-1} + \alpha\,(A - I_n)\Big]^{-1}. \tag{6}$$

It follows that the transformed phenotypes $\mathbf{z} = U^*\mathbf{y}$ are independent normal with mean 0 and variance

$$\mathrm{Var}(z_i) = \theta_0^{-1}\frac{1+\phi_0 s_i^2}{1-\alpha+\alpha s_i^2(1-\phi_0)+\alpha\phi_0 s_i^4}\,.$$

Suppose the data have come from this phenotypic–repulsion model, and we analyze them using the random-effects model. While it is always possible to get $\hat{\psi} < 0$ because of random fluctuations, we would like to show that the heritability implied by this model is "really" negative, in the sense that the distribution of $\hat{\psi}$ converges to a strictly negative value as the number of subjects goes to ∞. This will follow from Proposition 4.1 (below) with

$$f(t) = \frac{1+\phi_0 t}{1-\alpha+\alpha(1-\phi_0)t+\alpha\phi_0 t^2},$$

as long as $\phi_0 < \alpha$, since

$$f'(t) = \frac{\phi_0 - \alpha(1+\phi_0 t)^2}{(1-\alpha+\alpha(1-\phi_0)t+\alpha\phi_0 t^2)^2},$$

which is less than 0 for all $t \geq 0$.

In other words, to the extent that we say that heritability is defined by the linear model, heritability can be negative if genotypes and phenotypes interact through the environment in a manner like the phenotypic repulsion model. This will be true even if the phenotypic interactions are limited to small family groups. We prove that this is the case — that the heritability to which the estimates converge with increasing population size is negative — in the following Proposition.

Proposition 4.1. *Suppose we have a family of $n \times n$ GRMs A_n for $n \to \infty$, with eigenvalues $s_{n,i}^2$. We suppose that the distributions of eigenvalues converge to a nontrivial distribution $d\sigma(s^2)$.*

Let $U^{(n)}$ be the corresponding eigenvector matrix. For each n we have a multivariate normal random vector $\mathbf{y}^{(n)}$ with covariance matrix $U^{(n)}\,\mathrm{diag}\big(f(s_{n,i}^2)\big)U^{(n)*}$, where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a strictly decreasing, continuously differentiable function. We assume that the singular values $s_{n,i}$ are bounded above by $s_{\max}$, and that the following conditions hold:

$$C_1 := \inf_{0\leq t\leq s_{\max}^2} (-f'(t)), \tag{7}$$

$$C_2 := \inf_n \left(n^{-1}\sum_{i=1}^{n} s_{n,i}^4 - 1 \right) > 0. \tag{8}$$

(We maintain the normalization assumption that $\sum_i s_{n,i}^2 = n$.)

Let $\hat{\psi}_n$ be the MLE for an observation $\mathbf{y}^{(n)}$, calculated from the random-effects model with GRM A_n. Then $\hat{\psi}$ is bounded above in probability by a strictly negative quantity $-\delta$, depending on C_1 and σ, as $n \to \infty$. That is, the probability of $\hat{\psi}_n > -\delta$ goes to 0 as $n \to \infty$.

Proof. We follow the general principle, enunciated by [17], that the MLE for the misspecified model will converge to the closest fit in the Kullback–Leibler sense. In other words, the parameter estimate converges in probability to the location of the maximum *expected value* of the log-likelihood function. The arguments of [17] do not apply directly here, because we are not sampling i.i.d. random variables; however, the score function may be written

$$\frac{1}{\bar{v}(\psi)} \cdot g_n(\psi; \mathbf{x}) := \frac{1}{n(1-\psi)\bar{v}(\psi)} \sum_{i=1}^{n} a_i(\psi) x_i, \tag{9}$$

for $-1/(s_{\max}^2 - 1) < \psi \leq 1$, where (x_i) are i.i.d. χ_1^2 random variables and

$$a_i(\psi) := \frac{f(s_{n,i}^2)}{1 - \psi + \psi s_{n,i}^2} \left(\frac{s_{n,i}^2}{1 - \psi + \psi s_{n,i}^2} - n^{-1} \sum_{j=1}^{n} \frac{s_{n,j}^2}{1 - \psi + \psi s_{n,j}^2} \right). \tag{10}$$

We note that the maximum likelihood occurs either at a zero of g_n, or at $\psi = 1$ if g_n is everywhere positive. (It goes to $+\infty$ at the left boundary.) The coefficients $a_i(\psi)$ are uniformly bounded and uniformly Lipschitz, so, by the main theorem of [20], $g_n(\psi; \mathbf{x})$ converges uniformly in ψ to the function that is the limit of the expected values

$$G(\psi) = \lim_{n\to\infty} g_n(\psi; 1) = \frac{1}{1-\psi} \operatorname{Cov}_\sigma \left(\frac{f(S^2)}{1 - \psi + \psi S^2}, \frac{S^2}{1 - \psi + \psi S^2} \right).$$

(The covariance is understood here to be with respect to S^2 having distribution σ.) We need to show that G is negative for all ψ above the bound given in (11).

Near $\psi = 0$ the function $g_n(\psi; 1)$ is well-behaved, and takes on the value $\operatorname{Cov}(f(s_{n,i}^2), s_{n,i}^2)$ at $\psi = 0$. Since $f(t) + C_1 t$ is a decreasing function of t, for $t \in [0, s_{\max}^2]$, we have by the FKG inequality

$$\begin{aligned} G(0) &= \operatorname{Cov}\left(f(S^2) + C_1 S^2, S^2 \right) - C_1 \operatorname{Var}\left(S^2 \right) \\ &\leq -C_1 C_2 \\ &< 0. \end{aligned}$$

We also have

$$\begin{aligned} (1-\psi) G'(\psi) = &- \operatorname{Cov}\left(\frac{f(S^2)}{1 - \psi + \psi S^2}, \frac{S^2(S^2-1)}{(1 - \psi + \psi S^2)^2} \right) \\ &- \operatorname{Cov}\left(\frac{(S^2-1) f(S^2)}{(1 - \psi + \psi S^2)^2}, \frac{S^2}{1 - \psi + \psi S^2} \right) \\ &+ (1-\psi)^{-1} \operatorname{Cov}\left(\frac{f(S^2)}{1 - \psi + \psi S^2}, \frac{S^2}{1 - \psi + \psi S^2} \right) \end{aligned}$$

Since f is decreasing, we have for $0 > \psi > -\frac{1}{2s_{\max}^2 - 1}$ the bound

$$\left| G'(\psi) \right| \leq 20 s_{\max}^2 (s_{\max}^2 - 1) f(0).$$

Thus $G(\psi) < 0$ for all

$$\psi \in \left(-\frac{C_1 C_2}{20 s_{\max}^2 (s_{\max}^2 - 1) f(0)}, 0 \right). \tag{11}$$

It follows that $g_n(\psi; 1)$ is negative for all ψ between 0 and the bound given in (11), with probability tending to 1 as $n \to \infty$.

We note now that for $\psi \in [0, 1]$ $f(t)/(1 - \psi + \psi t)$ is a decreasing function of t, and $t/((1 - \psi + \psi t)$ is increasing, so (again by the FKG inequality) $G(\psi) < 0$, which completes the proof.

□

References

[1] Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins. *Behavior Genetics*, 43(5):415–426, 2013.

[2] A Earl Bell. Heritability in retrospect. *The Journal of Heredity*, 68(5):297–300, 1976.

[3] Gaurav Bhatia, Alexander Gusev, Po-Ru Loh, Bjarni J Vilhjálmsson, Stephan Ripke, Schizophrenia Working Group Psychiatric Genomics C, Shaun Purcell, Eli Stahl, Mark Daly, Teresa R de Candia, Kenneth S Kendler, Michael C O'Donovan, Sang Hong Lee, Naomi R Wray, Benjamin M Neale, Matthew C Keller, Noah A Zaitlen, Bogdan Pasaniuc, Jian Yang, and Alkes L Price. Haplotypes of common SNPs can explain missing heritability of complex diseases. *BioRxiv*, pages 1–25, July 2015.

[4] Guo-Bo Chen. Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman–Elston regression. *Frontiers in Genetics*, 5:107, 2014.

[5] Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, Stephan Ripke, Felix R Day, Shaun Purcell, Eli Stahl, Sara Lindström, John R B Perry, Yukinori Okada, Soumya Raychaudhuri, Mark J Daly, Nick Patterson, Benjamin M Neale, and Alkes L Price. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, September 2015.

[6] JL Gill and EL Jensen. Probability of obtaining negative estimates of heritability. *Biometrics*, pages 517–526, 1968.

[7] David Golan, Eric S Lander, and Saharon Rosset. Measuring missing heritability: inferring the contribution of common variants. *Proceedings of the National Academy of Sciences of the United States of America*, 111(49):E5272–81, December 2014.

[8] J. B. S. Haldane. The negative heritability of neonatal jaundice. *Annals of Human Genetics*, 60:3–5, 1996.

[9] Albert Jacquard. Heritability: one word, three concepts. *Biometrics*, pages 465–477, 1983.

[10] S. E. Johnston, D. Beraldi, A. F. McRae, J. M. Pemberton, and J. Slate. Horn type and horn length genes map to the same chromosomal region in Soay sheep. *Heredity*, 104:196–205, 2010.

[11] Thomas Liggett. *Stochastic Interacting Systems : Contact, Voter, and Exclusion Processes*. Springer Verlag, New York, 1999.

[12] Jay L Lush. Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. *Proceedings of the American Society of Animal Nutrition*, 1940(1):293–301, 1940.

[13] JA Nelder. The interpretation of negative components of variance. *Biometrika*, 41(3/4):544–548, 1954.

[14] David Steinsaltz, Andrew Dahl, and Kenneth W Wachter. Statistical properties of simple random-effects models for genetic heritability. bioRxiv: 087304, November 2016.

[15] Frank J Sulloway. Why siblings are like darwin's finches: Birth order, sibling competition, and adaptive divergence within the family. *The evolution of personality and individual differences*, pages 87–119, 2010.

[16] Campbell O Webb, David D Ackerly, Mark A McPeek, and Michael J Donoghue. Phylogenies and community ecology. *Annual review of ecology and systematics*, 33(1):475–505, 2002.

[17] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.

[18] Jian Yang, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, Andrew C Heath, Nicholas G Martin, Grant W Montgomery, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42(7):565–569, 2010.

[19] Jian Yang, Taeheon Lee, Jaemin Kim, Myeong-Chan Cho, Bok-Ghee Han, Jong-Young Lee, Hyun-Jeong Lee, Seoae Cho, and Heebal Kim. Ubiquitous polygenicity of human complex traits: genome-wide analysis of 49 traits in Koreans. *PLoS Genetics*, 9(3):e1003355, 2013.

[20] Ke-Hai Yuan. A theorem on uniform convergence of stochastic functions with applications. *journal of multivariate analysis*, 62(1):100–109, 1997.

[21] Zhihong Zhu, Andrew Bakshi, Anna A E Vinkhuyzen, Gibran Hemani, Sang Hong Lee, Ilja M Nolte, Jana V Van Vliet-Ostaptchouk, Harold Snieder, LifeLines Cohort Study, Tõnu Esko, Lili Milani, Reedik Mägi, Andres Metspalu, William G Hill, Bruce S Weir, Michael E Goddard, Peter M Visscher, and Jian Yang. Dominance genetic variation contributes little to the missing heritability for human complex traits. *American journal of human genetics*, 96(3):377–385, March 2015.