

Large effects and the infinitesimal model

Todd L. Parsons* & Peter L. Ralph†

December 3, 2023

Abstract

The infinitesimal model of quantitative genetics relies on the Central Limit Theorem to stipulate that under additive models of quantitative traits determined by many loci having similar effect size, the difference between an offspring’s genetic trait component and the average of their two parents’ genetic trait components is Normally distributed and independent of the parents’ values. Here, we investigate how the assumption of similar effect sizes affects the model: in particular, if the tail of the effect size distribution is polynomial with exponent $\alpha < 2$, then sums of effects should be well-approximated by a “stable distribution”, and find tail exponents between 1 and 2 in effect sizes estimated by genome-wide association studies of many human disease-related traits. We show that the independence of offspring trait deviations from parental averages in many cases implies a Gaussian distribution, suggesting that non-Gaussian models of trait evolution must explicitly track the underlying genetics, at least for loci of large effect. We also characterize possible limiting trait distributions of the infinitesimal model with infinitely divisible noise distributions, and compare our results to simulations.

1 Introduction

Much of the work to understand the evolution of quantitative traits in sexual species has focused on models in quantitative genetics that describe evolution of trait values without explicitly representing underlying genetic variation, but in a manner deriving from our understanding of how genetic variation works in practice. Such models are therefore interestingly intermediate between mechanistic models of genetics and purely phenomenological. One concrete model of heritable trait evolution is the *infinitesimal model*, which assumes a simple, explicit form for how the genetic component of offsprings’ traits are determined by their parents: they are Normally distributed, with the mean for each offspring equal to the average of the parents’ values, and a covariance independent of the parental trait values (but that depends on how they are related). ? give a history of the model and a formal proof of conditions under which this would be a good approximation to reality, which they extended to include dominance in ?.

As suggested by its name, under the infinitesimal model the effects of any given genetic variant on the trait is small (indeed, vanishing), and selection-driven trait shifts are mediated by small changes in frequencies of many alleles of small effect. However, empirical understanding of the genetic basis of trait variation in real-world species often finds a significant portion of heritable variation explained by variants at only a few loci. There are many possible explanations for this observation, including selection or publication bias, filters for larger-effect variants due to spatially varying selection [???], and/or grouping together of concordant alleles by linkage or inversions [?]. However, it seems clear that in at least those cases where we understand the underlying molecular mechanisms, mutations of large effect can be important in practice. The consensus in the field seems to be that traits exist on a spectrum between monogenic and polygenic, and that many - if not most - lie in the intermediate region, where heritable trait variation is due both to segregating large- and small-effect variants. Indeed, some sensible verbal models of the biological mechanisms by which a genetic change comes to have an effect on an organism’s trait describe a hierarchy or range of “proximities” to the

*LPSM, Sorbonne Université, CNRS UMR 8001, Paris, France, 75005

†Institute of Ecology & Evolution, University of Oregon, Eugene, OR, USA, 97405

trait, by which changes to more “nearby” genes have the possibility of causing larger changes in the trait [??]. This suggests a mixture model for the effect size distribution: perhaps the distribution of effect sizes within each gene is Normal, but with a standard deviation that varies with “proximity” to the trait; such mixtures can easily have large proportions of extreme values.

The Gaussian distribution appears thanks to “the” Central Limit Theorem, i.e., the universality of adding up a great many very small and independent things. However, there is a wider class of Central Limit Theorems, which describe the distributions of sums of a great many independent and exchangeable things: the *stable laws* [?]. Roughly speaking, these say that if we add together n independent copies of a random variable X for which $\mathbb{P}\{X > t\}$ is proportional to $t^{-\alpha}$ for some $0 < \alpha < 2$, then – almost regardless of the details – the result, scaled by $n^{1/\alpha}$, is well-approximated by a universal form, an α -stable distribution (and the approximation becomes exact, as $n \rightarrow \infty$). In such ensembles, single large entries can be important: concretely, the largest of the n copies will be of order $n^{1/\alpha}$, so that e.g., for $\alpha = 1$ a positive proportion of the sum of n entries will be contributed by a single one.

This suggests exploring whether stable distributions might reasonably stand in for the Gaussian in the infinitesimal model. In this paper, we make preliminary explorations in that direction: Do the results of ? carry over when the distribution of effect sizes falls in the domain of attraction of a stable law? The immediate answer to this will be “no” – in fact, we will see that independence of offsprings’ deviations and midparent values *implies* a Gaussian distribution in some sense. This independence is a key reason why the infinitesimal model is tractable, and so a conclusion is that additive models with non-Gaussian effects may not be well-approximated by “trait-only” models that do not explicitly represent underlying genetic variation. Having thus shown that “infinitesimal models with non-Gaussian noise” is on weak footing as a model of reality, we then go on to analyze exactly these models. (As motivation, we have not shown that such models are *not* good approximations to reality, and it is interesting and perhaps informative to understand how the noise distribution affects results.)

Stable distributions are by no means new to population biology. For instance, it is a long-standing question how often the motions of organisms are better modeled by essentially Brownian random walks or by Lévy processes [?] and a growing literature studies the effects of such power law distributions on spatial genetic patterns and evolutionary dynamics [e.g., ??]. More directly analogous to the problem we study here, ? found that an α -stable process (with $\alpha \approx 3/2$) fit the evolution of log endocranial volume on the primate phylogeny better than a Brownian model. It is also important to note that the infinitesimal model implies a Gaussian distribution for the difference between an offspring’s trait and their parents’ average, *not* for the entire population. The question of when the trait distribution in an evolving population might be modeled as a Gaussian has been discussed extensively [e.g., ??].

The ways in which genetic variants affect an organisms’ trait can be complex, and interactions between alleles, between loci, and with the environment are often important in practice. In fact, many traits claimed to be genetically determined in fact only appear heritable through non-genetic means, e.g., systematic inequalities in resource distribution partitioned by class or race. In this paper, we follow the literature in working with the relatively simple purely additive case of a trait that can be decomposed into a sum across loci of the contributions of the alleles from each locus, even omitting an independent effect of “the environment”. However, we emphasize that such models should not always be the default for understanding biological variation, especially for human traits.

Notation

We generally follow the notation of ?, although in a simpler situation, as we do not consider full pedigrees but rather only the haploid offspring of a single set of haploid parents (as produced, for instance, by a round of meiosis in a diploid stage), and ignore non-genetic contributions to the trait; so, below we say “trait” to mean “genetic contribution to the trait”.

Suppose that the focal individual’s trait is

$$Z_M = \sum_{\ell=1}^M \eta_{\ell} / M^{1/\alpha},$$

where η_ℓ is the contribution of the allele carried at the ℓ^{th} locus. The effects of the alleles carried by the two parents are $\eta_\ell^{[1]}$ and $\eta_\ell^{[2]}$, respectively, and their traits are $Z_M^{[1]} = \sum_\ell \eta_\ell^{[1]}/M^{1/\alpha}$ and $Z_M^{[2]} = \sum_\ell \eta_\ell^{[2]}/M^{1/\alpha}$. We assume the loci are unlinked and inheritance is fair, so letting X_ℓ be a family of Bernoulli(1/2) random variables that are independent of everything else,

$$\eta_\ell = X_\ell \eta_\ell^{[1]} + (1 - X_\ell) \eta_\ell^{[2]}.$$

What we will call the *Mendelian sampling term* (or occasionally *Mendelian noise*), R_M , is the offspring's trait minus the midparent value, \bar{Z}_M :

$$\begin{aligned} \bar{Z}_M &= \frac{Z_M^{[1]} + Z_M^{[2]}}{2} \\ R_M &= Z_M - \bar{Z}_M \\ &= \sum_{\ell=1}^M \frac{X_\ell(\eta_\ell^{[1]} - \eta_\ell^{[2]}) + (1 - X_\ell)(\eta_\ell^{[2]} - \eta_\ell^{[1]})}{2}. \end{aligned}$$

Note in particular that the contribution of locus ℓ to R_M is $\pm|\eta_\ell^{[1]} - \eta_\ell^{[2]}|$.

2 Empirical observations

It is mathematically intriguing to speculate what effect a power law effect size distribution for would have on the induced evolutionary dynamics. But, is there any evidence that non-Gaussian effects are important in practice? In this paper we focus primarily on the theory, but first take a brief and imperfect look at what some data have to say on the topic. To do this, we downloaded the GWAS results provided by ? for 7,221 phenotypes from 361,194 human individuals from the March 2018 release of the UK Biobank. After filtering (see details below) we were left with 695 binary illness-related phenotypes. Taken at face value, these provide for each phenotype a set of estimates of the additive effect of the alternative allele at a set of single nucleotide polymorphisms (SNPs) on the phenotype relative to the reference allele. The phenotypes we consider were all binary and coded as 0/1, so estimates could be thought of as additive adjustments to the risk (so, effects are *not* on a logit scale). These estimates were obtained as the coefficient for the SNP in a simple linear model fit to each phenotype for each SNP, with the first 20 principal components and all combinations of age, age squared, and inferred sex as covariates; we considered effects of those SNPs with a reported p -value less than 10^{-8} . There are a great many possible issues with these data; however, the results seem consistent across many SNPs and phenotypes, and we know of no reason that possible methodological artifacts (e.g., confounding by uncorrected variables) would specifically affect the decay of the tail of effect sizes.

Since the mendelian sampling term is a sum of terms of the form

$$(X_\ell - 1/2)\eta_\ell^{[1]} + (1/2 - X_\ell)\eta_\ell^{[2]},$$

where $\eta_\ell^{[i]}$ is the effect of the allele in parent i at locus ℓ , the appropriate central limit theorem for the sum is determined by the tail behavior of $|(X - 1/2)\eta| = |\eta|/2$, where X is Bernoulli(1/2) and η is the effect of a uniformly chosen allele at a uniformly chosen locus. So, we'd like to know if $\mathbb{P}\{|X| > 2t\}$ is proportional to $t^{-\alpha}$, i.e., the value of

$$\lim_{t \rightarrow \infty} \frac{-\log \mathbb{P}\{|X| > 2t\}}{\log t} = \alpha,$$

if the limit exists. Let p_ℓ be the minor allele frequency at the ℓ^{th} locus; since the data we have assigns effect 0 to the major allele, $\eta_\ell = 0$ with probability $1 - p_\ell$. Therefore, if we have estimated effects for M SNPs with minor allele frequency p_ℓ and effect size e_ℓ for the minor allele, for $1 \leq \ell \leq M$, then

$$\mathbb{P}\{|\eta| > 2t\} \approx \frac{1}{M} \sum_{\ell: |e_\ell| > 2t} p_\ell.$$

So, to obtain a per-trait estimate of α , we plot $\log \frac{1}{M} \sum_{\ell: |e_\ell| > t} p_\ell$ against $\log t$ for a sequence of values of t , and fit a linear model to the tail (which we take to be all values of t for which the number of SNPs with absolute effect above t is between 20 and 10% of all SNPs). Examples are shown in Figure 1.

We applied this procedure to results from models fit on both sexes, after removing phenotypes related to non-medical traits (e.g., employment, drug treatments, and dietary choices). This left us with 1,108 phenotypes for which we had enough data (e.g., enough significant SNPs) to estimate the tail exponent. We then restricted to phenotypes having non-missing numbers of “cases” and “controls” and a sample size of at least 300,000 (i.e., not more than 61,194 of 361,194 missing phenotypes), and removed the 65 remaining phenotypes that are common (more than 10,000 cases), as these often showed very different patterns (e.g., effect size distributions). This left us with 695 phenotypes; plots including the 413 phenotypes removed by these last steps are provided in Supplementary Figure S1.

Resulting values of the tail exponent α are shown in Figure 2. Estimated tail exponents range from about 1 to 2.5, and 28.9% had estimated values less than 1.5. A concern is that “noisier” phenotypes, for which estimated effect sizes are less reliable, might lead to larger estimated tail exponents; for this reason, we looked for an association between tail exponent and two proxies for power, number of SNPs (i.e., number of SNPs with p -value less than 10^{-8}) and number of cases. Interestingly, the relationship is nonmonotonic, with exponents closer to $\alpha = 2$ for phenotypes with around 1000 cases, and lower exponents for both rarer and more common phenotypes. (A similar pattern is seen for “number of SNPs”, but this may be due to association with power and hence number of cases.) This may indicate a statistical artifact, or it may be a result of biological differences in genetic architecture between more and less common phenotypes.

3 Independence and the Gaussian assumption

We know from ? that if the $\eta_\ell^{[i]}$ are independent with mean zero and $\sum_{\ell=1}^M \text{Var}[\eta_\ell]/M \rightarrow \sigma^2$, then with $\alpha = 2$, the traits $(Z_M, Z_M^{[1]}, Z_M^{[2]})$ converge in distribution to a multivariate Gaussian distribution $(Z, Z^{[1]}, Z^{[2]})$ for which $Z - (Z^{[1]} + Z^{[2]})/2$ is independent of $Z^{[1]}$ and $Z^{[2]}$.

Perhaps surprisingly, there is a converse – in at least some situations, independence of the Mendelian sampling term implies a Gaussian distribution. To see this, divide the parental loci into groups according to the results of inheritance, defining

$$\begin{aligned} U_M^{[1]} &= \sum_{\ell=1}^M X_\ell \eta_\ell^{[1]}, & V_M^{[1]} &= \sum_{\ell=1}^M (1 - X_\ell) \eta_\ell^{[1]}, \\ U_M^{[2]} &= \sum_{\ell=1}^M (1 - X_\ell) \eta_\ell^{[2]}, & V_M^{[2]} &= \sum_{\ell=1}^M X_\ell \eta_\ell^{[2]}, \end{aligned}$$

so that $Z_M^{[1]} = U_M^{[1]} + V_M^{[1]}$ and $Z_M^{[2]} = U_M^{[2]} + V_M^{[2]}$, while $Z_M = U_M^{[1]} + U_M^{[2]}$ is the transmitted portion of the parental genomes, and we can define $V_M = V_M^{[1]} + V_M^{[2]}$ to be the untransmitted portion. Furthermore,

$$\bar{Z}_M = \frac{Z_M + V_M}{2} \quad \text{and} \quad R_M = \frac{Z_M - V_M}{2}.$$

Now, note that if inheritance (the X_ℓ) is independent of the per-locus effects ($\eta_\ell^{[i]}$), the latter are all independent of each other, and all effects at a given locus are drawn from the same distribution, then the inherited allele, $\eta_\ell^{[1]} X_\ell + \eta_\ell^{[2]} (1 - X_\ell)$ is independent of (and has the same distribution as) the non-inherited allele, $\eta_\ell^{[1]} (1 - X_\ell) + \eta_\ell^{[2]} X_\ell$ – they are simply two independent draws from the distribution of effects at locus ℓ . Therefore, the inherited alleles that sum to Z_M are independent of the non-inherited alleles that sum to V_M .

The Kac-Bernstein theorem states that if two independent random variables have the property that their sum and difference are independent of each other, then the random variables have a joint Gaussian

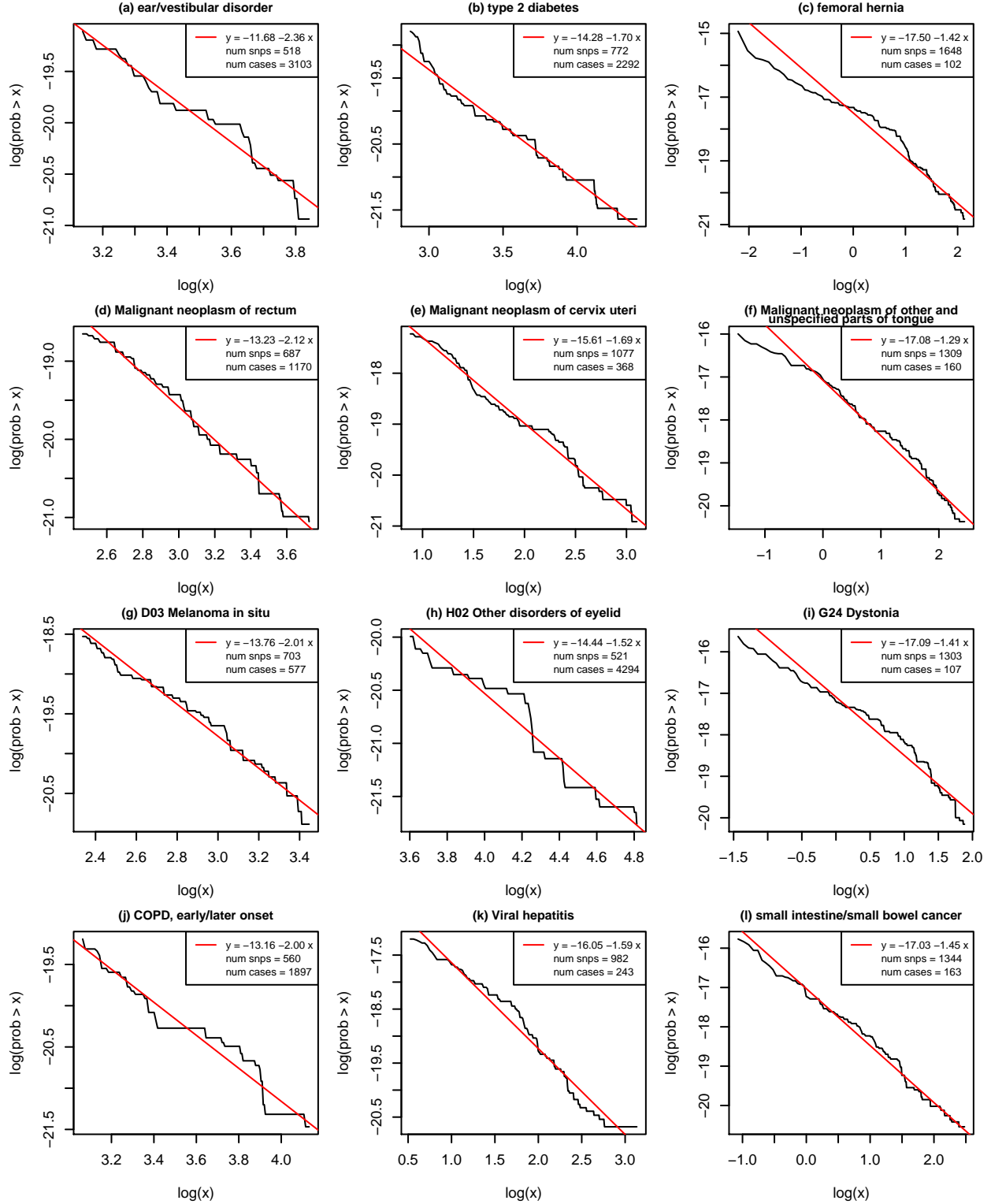


Figure 1: Tail distributions of frequency-weighted effects for twelve randomly chosen phenotypes. In each, for a phenotype with N SNPs, effect sizes e_i and frequencies p_i , the black line shows $\log \frac{1}{N} \sum_{i: |e_i| > t} p_i$ (vertical axis) against $\log t$ (horizontal axis); the red line shows the least-squares linear fit; each plot only covers the 10% largest-effect SNPs. Legends show the coefficients of the red line; the slope on ‘x’ is α , the estimated exponent; “num snps” is the number of SNPs for the phenotype with p -value less than 10^{-8} , and “num cases” is the reported number of the roughly 360,000 individuals recorded as having the listed phenotype.

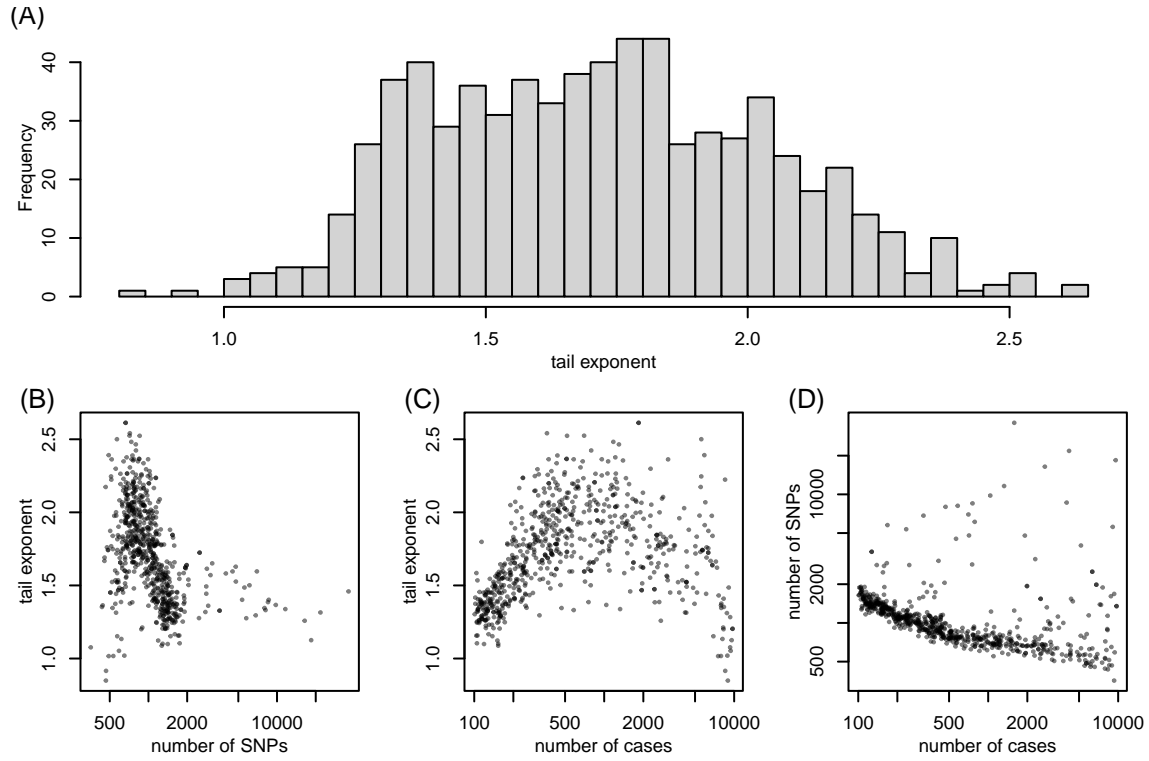


Figure 2: Estimated values of the tail exponent, α , for 775 illness-related binary phenotypes: (A) distribution of values; and plotted against (B) number of SNPs, and (C) number of cases.

distribution (??, and see ? for many much more general results). Since R_M is (half) the difference and \bar{Z}_M is (half) the sum of Z_M and V_M , and Z_M and V_M are independent, if we assume that also R_M and \bar{Z}_M are independent, then both are therefore jointly Gaussian.

We have shown the following:

Proposition 1. *Suppose that $\{\eta_\ell^{[i]}\}$ are independent for all ℓ and i , that the distribution of $\eta_\ell^{[i]}$ only depends on ℓ , and that $\{X_\ell\}$ are independent of $\{\eta_\ell^{[i]}\}$. If the Mendelian sampling term R_M is independent of the midparent value \bar{Z}_M , then R_M and \bar{Z}_M are jointly Gaussian.*

Since models can be easily set up for which trait distributions are not Gaussian, the implication of this is that for such models, independence of the Mendelian sampling term is not likely a good assumption. Note that the proof does not depend on independence of the X_ℓ : indeed, alleles may be inherited together, although in practice this would quickly create dependencies between nearby allelic effects. Also note that it is *not* usually true that the components – for instance, $U_M^{[1]}$ and $U_M^{[2]}$ – are independent of each other.

As we discuss more later, the presence of a large effect polymorphic allele makes R and \bar{Z} dependent: consider the extreme case where $\eta_\ell^{[i]}$ are i.i.d. with variance 1 for $\ell \geq 2$, while $\eta_1^{[i]} = \pm 10$ with probability 1/2 each, and $Z^{[i]} = \eta_1^{[i]} + \sum_{\ell=2}^M \eta_\ell^{[i]} / \sqrt{M}$. Then in the limit R depends strongly on parental trait values: if $Z^{[1]} \approx Z^{[2]}$ and so $\bar{Z} \approx \pm 10$ then most likely $\eta_1^{[1]} = \eta_1^{[2]}$ and so R is Normal with variance 1/2; but if $\eta_1^{[1]} \neq \eta_1^{[2]}$ (and so $\bar{Z} \approx 0$) then R is close to +5 or -5 with equal probability.

4 Abnormal quantitative genetics?

As observed in ?, there are at least three interpretations of what is called “the infinitesimal model”. The maximalist of these, what ? calls the Gaussian-population approximation, assumes a large number of small-effect loci, parental trait-independent Gaussian Mendelian sampling terms (as rigorously justified in ??), and, moreover, that population trait values are in deterministic mutation-selection balance, with a Gaussian empirical distribution (*i.e.*, the *fraction* of the population with trait value in $[z, z + dz]$ is approximately $\frac{1}{2\pi\sigma^2} e^{-(z-m)^2/(2\sigma^2)} dz$ for suitable choices of the empirical mean and variance m and σ^2). Leaving aside the question of whether or not some combination of effect sizes, dominance, and epistasis can justify non-Gaussian, trait-independent, Mendelian sampling terms, we might reasonably ask when (or if) deterministic mutation-selection balance is possible in a non-Gaussian model that retains independence of the Mendelian sampling terms. (Indeed, above we have put substantial restrictions on what models of inheritance might lead to such models.) We might also ask whether such a model might still reasonably approximate an additive model with non-Gaussian effects, and for the purpose of investigating this with simulations, some calculations will be helpful.

Concretely, consider a Moran process where an individual whose trait value is Z dies at rate $\mu(Z)$ and is replaced by the offspring of two randomly chosen individuals, *i.e.*, a new individual with trait

$$\frac{Z^1 + Z^2}{2} + R,$$

where Z^1 and Z^2 are independent draws from the empirical population distribution and R is an independent draw from the Mendelian sampling distribution (independent of Z^1 and Z^2 , with distribution to be specified). Mutation-selection balance demands that that the distribution of individuals removed by death and the distribution of individuals added by birth are equal, *i.e.*, that $(Z^1 + Z^2)/2 + R$ is equal in distribution to a μ -weighted choice of Z . Taking the characteristic function of both these distributions, deterministic mutation-selection balance is therefore equivalent to

$$\frac{\mathbb{E} \left[e^{i\theta Z} \mu(Z) \right]}{\mathbb{E}[\mu(Z)]} = \mathbb{E} \left[e^{i\theta \left(\frac{Z^1 + Z^2}{2} + R \right)} \right]. \quad (1)$$

Now, we assumed that Z^1 , Z^2 , and R are independent, so we can write (1) as

$$\begin{aligned}\frac{\mathbb{E}[e^{i\theta Z} \mu(Z)]}{\mathbb{E}[\mu(Z)]} &= \mathbb{E}\left[e^{i\theta\left(\frac{Z^1+Z^2}{2}+R\right)}\right] \\ &= \mathbb{E}\left[e^{i\frac{\theta}{2}Z^1}\right] \mathbb{E}\left[e^{i\frac{\theta}{2}Z^2}\right] \mathbb{E}\left[e^{i\theta R}\right] \\ &= \phi\left(\frac{\theta}{2}\right)^2 \Phi(\theta),\end{aligned}$$

where $\phi(\theta) = \mathbb{E}[e^{i\theta Z^1}] = \mathbb{E}[e^{i\theta Z^2}]$ and $\Phi(\theta) = \mathbb{E}[e^{i\theta R}]$ are the characteristic functions of Z^j and R , respectively. In what follows, we will explore some non-Gaussian solutions $\phi(\theta)$ to (1).

4.1 The Neutral Case

First, consider the case without selection, so μ is a constant which we may take equal to 1. Then, our condition on the characteristic function simplifies to

$$\phi\left(\frac{\theta}{2}\right)^2 \Phi(\theta) = \phi(\theta), \quad (2)$$

i.e., $\phi(\theta)$ is a fixed point of F , where

$$F(\phi)(\theta) = \phi\left(\frac{\theta}{2}\right)^2 \Phi(\theta),$$

whereas any fixed point must be the limit of an arbitrary characteristic function, $\varphi(\theta)$, under repeated applications of F , *i.e.*,

$$\begin{aligned}\phi(\theta) &= \lim_{n \rightarrow \infty} F^{(n)}(\varphi)(\theta) \\ &= \lim_{n \rightarrow \infty} \varphi\left(\frac{\theta}{2^n}\right)^{2^n} \prod_{k=1}^{n-1} \Phi\left(\frac{\theta}{2^k}\right)^{2^k}.\end{aligned}$$

is a fixed point whenever this limit exists.

Since φ and Φ are characteristic functions, they are continuous and $\phi(0) = \Phi(0) = 1$, and thus both are non-zero in a neighbourhood of 0. In particular, $\psi(\theta) = \text{Log } \varphi(\theta)$ and $\Psi(\theta) = \text{Log } \Phi(\theta)$ exist in some neighbourhood of 0 as well¹. From the above, we have

$$\text{Log } F^{(n)}(\varphi)(\theta) = 2^n \psi\left(\frac{\theta}{2^n}\right) + \sum_{k=1}^{n-1} 2^k \Psi\left(\frac{\theta}{2^k}\right).$$

We briefly explore what possible limiting values are valid characteristic functions. We refer to the first term on the left hand side as the *reproduction term* and the remaining sum as the *noise term*, and only consider to the case when both have well-posed limits independent of the other.

Noise Terms First, consider the noise terms. A necessary condition for $\sum_{k=1}^{\infty} 2^k \Psi\left(\frac{\theta}{2^k}\right)$ to converge is that $2^k \Psi\left(\frac{\theta}{2^k}\right) \rightarrow 0$; since $\Psi(0) = 0$, this implies that if it exists, $\Psi'(0+) = 0$. In particular, as we would expect, if R has a mean, it must be 0. If, on the other hand, $\Psi(\theta)$ is twice differentiable, so R has mean 0 and finite variance, say $\Psi''(0) = \eta^2$, then Taylor's theorem tells us that for θ sufficiently small,

$$\Psi(\theta) = \frac{\eta^2}{2} \theta^2 + r(\theta) \theta^2,$$

¹We use Log to denote the principal value of the complex logarithm, for which the imaginary part always lies in the interval $(-\pi, \pi]$.

where $r(\theta) \rightarrow 0$ as $\theta \rightarrow 0$. Thus,

$$\frac{2^{k+1}\Psi\left(\frac{\theta}{2^{k+1}}\right)}{2^k\Psi\left(\frac{\theta}{2^k}\right)} = \frac{1}{2} \frac{\eta^2\theta^2 + r\left(\frac{\theta}{2^{k+1}}\right)\theta^2}{\eta^2\theta^2 + r\left(\frac{\theta}{2^k}\right)\theta^2} \rightarrow \frac{1}{2},$$

as $k \rightarrow \infty$ and the sum converges by the ratio test. In this case,

$$\begin{aligned} \sum_{k=1}^{n-1} 2^k \Psi\left(\frac{\theta}{2^k}\right) &= \sum_{k=1}^{n-1} 2^k \left(-\frac{\eta^2}{2} \frac{\theta^2}{2^{2k}} + r\left(\frac{\theta}{2^k}\right) \frac{\theta^2}{2^{2k}} \right) \\ &= -\eta^2 \theta^2 (1 - 2^{-n+1}) + \sum_{k=1}^{n-1} r\left(\frac{\theta}{2^k}\right) \frac{\theta^2}{2^k} \end{aligned}$$

If we assume that $r(\theta) \equiv 0$, so that the noise is $N(0, \eta^2)$, then the limiting distribution is $N(0, 2\eta^2)$. When $r(\theta)$ is non-zero, we don't expect the higher order terms to vanish.

It is unclear what more can be said in general. If however, we specialize to the case when R is a stable law, we can explicitly compute the limiting noise term. We recall that if $\Psi(\theta)$ corresponds to a stable law (see *e.g.*, ?), then

$$\Psi(\theta) = im\theta + c|\theta|^\alpha (1 - i\beta \operatorname{sgn}(\theta)\omega_\alpha(\theta))$$

for $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $c \geq 0$, and $m \in \mathbb{R}$ and

$$\omega_\alpha(\theta) = \begin{cases} \tan \frac{\pi\alpha}{2} & \text{if } \alpha \neq 1, \text{ and} \\ -\frac{2}{\pi} \log |\theta| & \text{if } \alpha = 1. \end{cases}$$

By our general remarks above, we can assume $m = 0$ and exclude the case $\alpha = 1$, $\beta = 0$, for which $\Psi'(0+) = c$. We then have

$$\begin{aligned} \sum_{k=1}^{n-1} 2^k \Psi\left(\frac{\theta}{2^k}\right) &= \sum_{k=1}^{n-1} 2^{(1-\alpha)k} c |\theta|^\alpha (1 - i\beta \operatorname{sgn}(\theta) \left(\frac{\theta}{2^k}\right) \omega_\alpha\left(\frac{\theta}{2^k}\right)) \\ &= \begin{cases} c |\theta|^\alpha \frac{1 - 2^{(1-\alpha)n}}{1 - 2^{1-\alpha}} (1 - i\beta \operatorname{sgn}(\theta) \tan \frac{\pi\alpha}{2}) & \text{if } \alpha \neq 1, \text{ and} \\ (n-1) c |\theta|^\alpha \left(1 + \frac{2i\beta}{\pi} \operatorname{sgn}(\theta) (\log |\theta| - \frac{n}{2} \log 2)\right) & \text{if } \alpha = 1, \end{cases} \end{aligned}$$

which converges as $n \rightarrow \infty$ if and only if $\alpha > 1$, in which case the limit is

$$\frac{c}{1 - 2^{1-\alpha}} |\theta|^\alpha \left(1 - i\beta \operatorname{sgn}(\theta) \tan \frac{\pi\alpha}{2}\right), \quad (3)$$

which corresponds to a $(\alpha, \beta, \frac{c}{1-2^{1-\alpha}}, 0)$ -stable law.

Reproductive Terms: We now turn our attention to the limit

$$\alpha(\theta) = \lim_{n \rightarrow \infty} 2^n \psi\left(\frac{\theta}{2^n}\right)$$

If it exists, then for all integers k ,

$$\alpha(2^k \theta) = \lim_{n \rightarrow \infty} 2^n \psi\left(\frac{2^k \theta}{2^n}\right) = \lim_{n \rightarrow \infty} 2^k 2^{n-k} \psi\left(\frac{\theta}{2^{n-k}}\right) = 2^k \alpha(\theta).$$

and, since $\alpha(0) = 0$,

$$\frac{\alpha(\theta)}{\theta} = \frac{\alpha(2^k \theta)}{2^k \theta} \rightarrow \begin{cases} \alpha'(0+) & \text{if } \theta > 0, \text{ and} \\ \alpha'(0-) & \text{if } \theta < 0, \end{cases}$$

as $k \rightarrow -\infty$ if the limits exist (*i.e.*, there are unique left-and right-hand limits at zero; this is the case if *e.g.*, we require $\alpha(\theta)$, and thus $\phi(\theta)$, to be convex functions). We henceforth assume this.

Since $\varphi(\theta)$ is a characteristic function, it has the Hermitian property: $\phi(-\theta) = \overline{\phi(\theta)}$, and thus $\psi(-\theta) = \overline{\psi(\theta)}$ and $\alpha(-\theta) = \overline{\alpha(\theta)}$. In particular, if one (and by extension both) limits exist, we must have $\alpha'(0-) = \alpha'(0+)$. In this case, setting $-\gamma + im = \alpha'(0+)$, we have

$$\alpha(\theta) = \begin{cases} im\theta - \gamma\theta & \text{if } \theta \geq 0, \text{ and} \\ im\theta + \gamma\theta & \text{if } \theta < 0. \end{cases}$$

If $\gamma > 0$, then we recognize $e^{\alpha(\theta)} = e^{im\theta - \gamma|\theta|}$ as the characteristic function of a Cauchy distribution with median 0 and median absolute deviation γ , whereas if $\gamma = 0$, then $e^{\alpha(\theta)} = e^{im\theta}$ is the Fourier transform of a Dirac mass at m , *i.e.*, the characteristic function for a monomorphic population with all individuals having trait value equal to m . (And, $\gamma < 0$ does not give a valid characteristic function.)

Blending inheritance redux. Note that the characteristic functions constructed above are solutions to (2) when transmission is noiseless, and the offspring have exactly the parental mean trait – *i.e.*, blending inheritance. The solution $\phi(\theta) = e^{i\theta m}$ is exactly the reversion to the mean that Darwin’s detractors used to argue against the theory of evolution [??]. The solution $\phi(\theta) = e^{im\theta - \gamma|\theta|}$ offers an amusing counter-factual: if a Lamarkian population could somehow achieve a Cauchy-distributed trait distribution, it would maintain that diversity indefinitely.

More generally, one will get convergence if the noise is a mixture of components with finite variance and stable laws as above, whereas the population distribution will be a mixture of a (possibly degenerate) Cauchy distribution and such noise terms. Note that in the presence of noise, a pure Cauchy-distributed solution is *not* possible in the neutral case.

4.2 Stabilizing Selection

Unfortunately, once selection is included, (1) is less amenable to the sort of simple arguments above. We will content ourselves with presenting two examples of stabilizing selection, one already familiar to quantitative genetics, the other a Cauchy solution obtained by trial-and-error, which illustrates how selection can stabilize an otherwise unviable solution.

Gaussian Solution: First, consider the case when $\mu(x) = e^{ax^2}$. In this case there is a Gaussian solution: suppose Z is Gaussian with mean m and variance σ^2 , so $\phi(\theta) = \exp(im\theta - \frac{1}{2}\sigma^2\theta^2)$. Moreover, provided $a < \frac{1}{2\sigma^2}$, we can explicitly compute $\frac{\mathbb{E}[e^{i\theta Z} \mu(Z)]}{\mathbb{E}[\mu(Z)]}$: if R is Gaussian with mean zero and variance η^2 , then equation (1) becomes

$$e^{im\theta - \frac{1}{2}(\frac{\sigma^2}{2} + \eta^2)\theta^2} = e^{i\frac{m}{1-2a\sigma^2}\theta - \frac{1}{2}\frac{\sigma^2}{1-2a\sigma^2}\theta^2},$$

which, equating real and complex components, has a solution provided $m = 0$ and

$$\sigma^2 = \frac{\sqrt{16a^2\eta^4 + 24a\eta^2 + 1} - 4a\eta^2 - 1}{4a} = 2\eta^2(1 - 4a\eta^2) + O(a^2).$$

Note that we do not require very weak selection to guarantee a Gaussian steady state, only that $a < \frac{1}{4\eta^2}$. However, this does not imply that the transient distribution is also Gaussian, as is sometimes assumed [*e.g.*, ?].

4.2.1 Cauchy Solution

Almost uniquely among the stable laws, the Cauchy distribution has a probability density function expressible in elementary functions: for a Cauchy random variable with median (location) x_0 and mean absolute

deviation (scale) γ , the density is

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]}. \quad (4)$$

Armed with this density function, we can explicitly compute the right hand side of (1), assuming that

- (i) the population is in mutation-selection balance corresponding to a Cauchy density with median 0 and scale $\gamma > 0$,
- (ii) selection acts as $\mu(x) = \frac{\delta^2}{\gamma^2} \frac{\gamma^2 + x^2}{\delta^2 + x^2}$ for some $\delta > \gamma > 0$, and
- (iii) the Mendelian sampling term R is given by a Cauchy(0, $\delta - \gamma$) random variable (*i.e.*, corresponding to a stable (1, 0, $\delta - \gamma$, 0) law).

Under these assumptions, (1) is

$$\begin{aligned} \frac{\mathbb{E}[e^{i\theta Z} \mu(Z)]}{\mathbb{E}[\mu(Z)]} &= e^{-\delta|\theta|} \\ &= e^{-2\gamma|\frac{\theta}{2}| - (\delta - \gamma)|\theta|}, \end{aligned}$$

which we recognize as being equal to $\phi\left(\frac{\theta}{2}\right)^2 \Phi(\theta)$, where $\phi(\theta) = e^{i\gamma|\theta|}$ and $\Phi(\theta) = e^{i(\delta - \gamma)|\theta|}$ are the characteristic functions of the population density and the Mendelian sampling term, respectively, giving a non-Gaussian solution to (1) with both noise and selection.

As remarked above, we found this solution by inspection and trial-and-error; it is interesting to ask if any similar examples can be constructed for other stable laws. The lack of an elementary probability density function unfortunately rules out the naïve approach used here.

5 The difficulty with an abnormal quantitative genetics

Above we have seen that (a) in principle, a mutation-selection balance is possible with a Cauchy empirical distribution and Cauchy Mendelian noise, but (b) such a model cannot derive from a simple model of additive loci. To that end, it's instructive to see where the argument of ? breaks down if instead of assuming bounded allele effects of order $\mathcal{O}(1/\sqrt{M})$, we assume that the allele at locus ℓ in individual j contributes $\eta_\ell^{[j]}/M$, where the $\eta_\ell^{[j]}$ are independent Cauchy(0, γ_ℓ) random variables. In this case, the typical effect size at locus ℓ is of order $\frac{\gamma_\ell}{M}$, which is asymptotically smaller than the allele effects assumed in ?, but large effect alleles are more likely.

The crucial step in the derivation of the infinitesimal model in ? is to show that, following the arguments in ?, knowing an individual's trait value gives vanishingly little information about the contribution of any given locus as $M \rightarrow \infty$, which in turn allows one to show that the within-family Mendelian segregation variance is independent of the parental traits. Here, we will consider how conditioning an individual's trait value affects the law of a single allelic contribution in the Cauchy setting described above. A similar argument can be made for any α -stable distribution, but the lack of explicit probability density functions for generic $\alpha \in (0, 2)$ adds complexity to the argument without adding any insight.

Consider the simplest possible case of an ancestral population, of N individuals, where the j^{th} individual has trait value

$$Z_M^{[j]} = \frac{1}{M} \sum_{\ell=1}^M \eta_\ell^{[j]},$$

where the $\eta_\ell^{[j]}$ are independent across individuals and across loci. By the additive property of Cauchy random variables, $Z_M^{[j]} \stackrel{i.i.d}{\sim} \text{Cauchy}(0, \bar{\gamma})$, where

$$\bar{\gamma} = \frac{1}{M} \sum_{\ell=1}^M \gamma_\ell$$

is the mean of the scale parameters.

To see what knowing the parental trait value $Z_M^{[j]}$ tells us about the contribution of any given locus, $\eta_\ell^{[j]}$, we follow the arguments in ? and ? using Bayes' Theorem. Proceeding informally, with the probabilities indicating probability density functions, we have

$$\begin{aligned} \mathbb{P}\{\eta_\ell^{[j]} = y | Z^{[j]} = z\} &= \frac{\mathbb{P}\{\eta_\ell^{[j]} = y; Z^{[j]} = z\}}{\mathbb{P}\{Z^{[j]} = z\}} \\ &= \frac{\mathbb{P}\{Z^{[j]} = z | \eta_\ell^{[j]} = y\}}{\mathbb{P}\{Z^{[j]} = z\}} \mathbb{P}\{\eta_\ell^{[j]} = y\} \\ &= \frac{\mathbb{P}\{Z^{[j]} - \frac{1}{M}\eta_\ell^{[j]} = z - \frac{1}{M}y\}}{\mathbb{P}\{Z^{[j]} = z\}} \mathbb{P}\{\eta_\ell^{[j]} = y\} \end{aligned}$$

Now, $Z^{[j]} - \frac{1}{M}\eta_\ell^{[j]} = \frac{1}{M} \sum_{k=1: k \neq \ell}^M \eta_k^{[j]}$ is Cauchy-distributed with scale $\bar{\gamma} - \frac{\gamma_\ell}{M}$. Using the probability density function for a Cauchy distributed random variable (4), a straightforward computation gives

$$\begin{aligned} \mathbb{P}\{\eta_\ell^{[j]} = y | Z^{[j]} = z\} &= \frac{f\left(z - \frac{y}{M}; 0, \bar{\gamma} - \frac{\gamma_\ell}{M}\right)}{f(z; 0, \bar{\gamma})} \mathbb{P}\{\eta_\ell^{[j]} = y\} \\ &= \left(1 + \frac{\gamma_\ell(\bar{\gamma}^2 - z^2) + 2zy}{\gamma_\ell(\bar{\gamma}^2 + z^2)} \frac{1}{M} + \mathcal{O}\left(\frac{1}{M^2}\right)\right) \mathbb{P}\{\eta_\ell^{[j]} = y\}. \end{aligned}$$

Noting that $|(\bar{\gamma}^2 - z^2)/(\bar{\gamma}^2 + z^2)| \leq 1$ and $|z/(\bar{\gamma}^2 + z^2)| \leq 1/(2\bar{\gamma})$ (since it is maximized at $z = \bar{\gamma}$), so

$$\left| \frac{\gamma_\ell(\bar{\gamma}^2 - z^2) + 2zy}{\gamma_\ell(\bar{\gamma}^2 + z^2)} \right| \leq \frac{|y|}{\gamma_\ell \bar{\gamma}},$$

we have that

$$\left| \mathbb{P}\{\eta_\ell^{[j]} = y | Z^{[j]} = z\} - \mathbb{P}\{\eta_\ell^{[j]} = y\} \right| \leq \left(\left(1 + \frac{|y|}{\gamma_\ell \bar{\gamma}}\right) \frac{1}{M} + \mathcal{O}\left(\frac{1}{M^2}\right) \right) \mathbb{P}\{\eta_\ell^{[j]} = y\}.$$

In the case of bounded allelic effects, the error introduced by approximating the effect-size distribution conditional on the trait value by the unconditioned distribution is proportional to $|Z^{[j]}|$, so that extreme trait values are most informative of the size of allelic effects, whereas here, the bound is independent of the trait value, and the error is maximized when the trait value is approximately equal to the mean absolute deviation of its distribution, *i.e.*, the median value of $|Z^{[j]}|$.

The argument above shows for a typical locus, the effect of conditioning on the trait gives negligible information about the contribution from that site. Recalling that

$$\mathbb{P}\left\{|\eta_\ell^{[j]}| > y\right\} = 1 - \frac{2}{\pi} \arctan\left(\frac{y}{\gamma_\ell}\right) \sim \frac{2\gamma_\ell}{\pi y},$$

we see that for any $\epsilon > 0$, and large values of M ,

$$\mathbb{P}\left\{\frac{\eta_\ell^{[j]}}{\gamma_\ell \bar{\gamma} M} > \epsilon\right\} \sim \frac{2\gamma_\ell}{\pi \gamma_\ell \bar{\gamma} M \epsilon}.$$

so the probability the error is larger than ϵ at any given locus decays rapidly as $M \rightarrow \infty$.

If, however, we consider not any one locus, but rather all loci, we get a different picture. The variables $\eta_\ell^{[j]}$ are independent, and thus so are the Bernoulli random variables

$$\mathbb{1}_\ell^M = \begin{cases} 1 & \text{if } \frac{\eta_\ell^{[j]}}{\gamma M} > \epsilon, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Now, since

$$\lim_{M \rightarrow \infty} \sum_{\ell=1}^M \frac{2\gamma_\ell}{\pi\gamma M\epsilon} = \frac{2}{\pi\epsilon},$$

the number of loci at which the magnitude of the effect size exceeds $M\epsilon$, which is $\sum_{\ell=1}^M \mathbb{1}_\ell^M$, converges weakly as $M \rightarrow \infty$ to a Poisson random variable with mean $2/(\pi\epsilon)$ (see *e.g.*, Theorem 2.6.1 in ?). In particular, for large values of M and fixed $\epsilon > 0$,

$$\mathbb{P} \left\{ \max_{i=1}^M \frac{\eta_\ell^{[j]}}{\gamma M} > \epsilon \right\} \sim 1 - e^{-\frac{2}{\pi\gamma\epsilon}},$$

which approaches 1 as $\epsilon \rightarrow 0$, while the expected number of M alleles such that $\frac{\eta_\ell^{[j]}}{\gamma M} > \epsilon$ is asymptotically $\frac{2}{\pi\epsilon}$.

In particular, there is at least one locus at which knowing the parental trait gives one information about the effect size (and there are on the order of $\mathcal{O}(\frac{1}{\epsilon})$ that give “an ϵ of information”), and these large effect loci in turn give rise to correlations between parental traits and the deviation of the offsprings’ traits from the midparent (i.e., the Mendelian sampling terms). However, the first calculations above suggest that for “most” loci, this is not true, and so perhaps a reasonable picture of Mendelian sampling terms in an additive model with Cauchy-distributed effects might be “some large Mendelian alleles plus Gaussian noise”. We now turn to simulation, to investigate.

6 Simulations

In simulation we can easily include realistic recombination (instead of unlinked loci) and selection. We used SLiM v4 [?] to simulate a discrete-time approximation to the continuous-time Moran model described above as follows. Individuals are diploid, hermaphroditic, and sexual; the genome is of length 10^8 bp with a uniform recombination rate of 10^{-8} crossovers per bp; mutations occur at a rate of 10^{-9} per bp. Each new mutation is independently assigned an “effect” drawn from the “effect size distribution” (and will be either Gaussian or Cauchy). The trait of each individual is determined additively by summing the effects of all mutations they carry across both chromosomes. Let $\mu(z)$ be the death rate of an individual with trait z , and let dt be a small constant (we took $dt = 0.01$). In all plots, one time unit is approximately one generation. Then, at each time step each individual dies with probability $1 - \exp(-dt\mu(z))$; if there are k such individuals chosen to die in a given time step, then we choose k individuals uniformly and without replacement to produce new offspring; for each such reproduction event the parent chooses a mate uniformly at random from the population. This maintains the population at a fixed size, N . (Note that individuals may self and individuals chosen to die may also reproduce, but in a large population these details should be unimportant.)

We first consider the neutral case (i.e., with $\mu \equiv 1$). As shown in Figure 3, a simulation with a Cauchy effect size distribution, unsurprisingly, has a distribution of trait values with more extreme values and a rougher path of population median trait value over time than does a simulation with a Gaussian effect size distribution. Indeed, the sum of all effects along the path from the start of the simulation to an arbitrary individual will approximate either Brownian motion (in the Gaussian case) or a Cauchy process (in the Cauchy case). Figure 4 shows how the Mendelian sampling term depends on midparent value. In the Gaussian case, the Mendelian noise is Gaussian and independent of midparent value, as shown by the

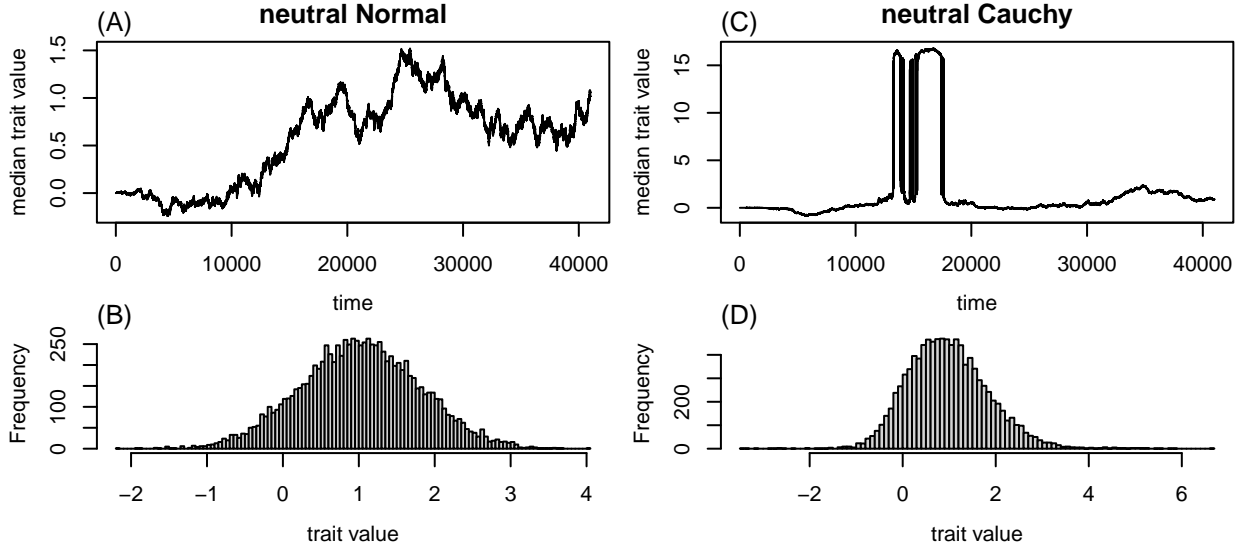


Figure 3: Median trait values over time (**A** and **C**) and distributions of final trait values (**B** and **D**) in neutral simulations with Gaussian (**A** and **B**) and Cauchy (**C** and **D**) effect size distributions. Simulations were begun with no genetic variation, had populations of size $N = 1000$, a genome of length $L = 10^8$ bp and mutation rate of $\mu = 10^{-9}$ per bp per generation; the effect size distributions were Normal with mean 0 and standard deviation $1/\sqrt{4NL\mu}$ and Cauchy with center 0 and scale $1/(4NL\mu)$, respectively.

horizontal lines at 1.0 in Figure 4C. On the other hand, in the Cauchy case we can see from either Figure 4D or F that Mendelian noise depends strongly on midparent trait, as we would expect in the presence of several large-effect alleles.

We next compared two situations of stabilizing selection, motivated by the results in Section 4.2: (1) the effect size distribution was Gaussian with mean zero and standard deviation 1.0 and $\mu(z) = e^{\alpha z^2}$; and (2) the effect size distribution was Cauchy, centered at zero and with median absolute deviation $\delta - \gamma$ and $\mu(z) = (\delta/\gamma)(\gamma + z^2)/(\delta + z^2)$ with $\delta = 0.5$ and $\gamma = 0.2$. Figures 5 and 6 show that the picture is largely unchanged by stabilizing selection except that, unsurprisingly, the median trait no longer wanders unconstrained (but still shows substantially larger fluctuations under the Cauchy model).

However, the two models look much more similar when considering only alleles of small effect. Figure 7 shows that midparent values and Mendelian sampling terms computed using only mutations with effect size less than 0.2 (excluding less than 1% of mutations) are quite close to independent. The result does not depend strongly on the cutoff chosen, and are similar even if less 0.1% of mutations are excluded.

7 Discussion

A substantial challenge for the field is to develop tractable models of trait evolution that include contributions of both “large” and “small” effect loci. In practice, breeders have long embellished the “animal model” (essentially, the infinitesimal model) to predict traits by combining both effects of known large effect loci and cumulative effects of many unknown loci as mediated by a relatedness matrix [e.g., ???]. We can study any model with simulation, but the way forwards to theoretical understanding seems murky.

Further analysis of the infinitesimal model with other noise distributions, as we have begun here, could be interesting, e.g., to see how changing the distribution of the Mendelian sampling term affects the rate of adaptive evolution, levels of genetic load, the rate of fixation, and the strength of linked selection. However,

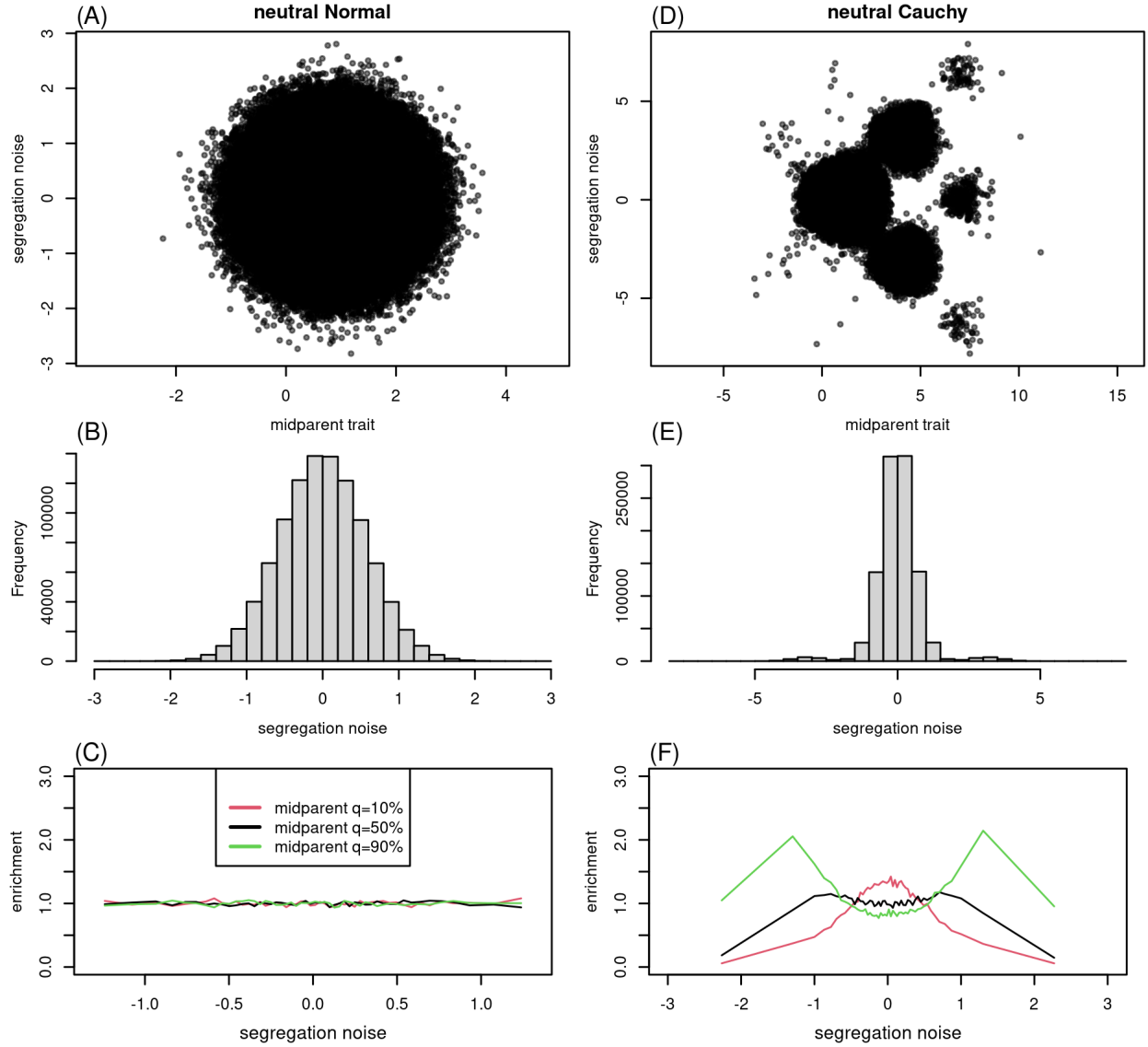


Figure 4: The distribution of Mendelian sampling terms with Gaussian **(A-C)** and Cauchy **(D-F)** effect size distributions, from the same simulations shown in Figure 3, for all individuals born during the last 1,000 time units of the simulation. **(A,D)** Mendelian noise (offspring trait minus midparent) plotted against midparent. **(B,E)** Histograms of the Mendelian sampling term. **(B,E)** Enrichment in quantiles of the Mendelian sampling term conditional on midparent value. To compute the latter, let $q_0, q_2, q_4, \dots, q_{98}, q_{100}$ divide the distribution of the Mendelian sampling term into 50 bins with roughly equal numbers in each (i.e., the quantiles), let $p_k(a, b)$ denote the proportion of the offspring of parents with midparent value is in $[a, b]$ whose Mendelian sampling term is in $[q_{k-1}, q_k)$, and let p_k be the proportion of all offspring with Mendelian noise in $[q_{k-1}, q_k)$ (so, $p_k \approx 0.02$). Then, each line shows $p_k(a, b)/p_k$ plotted against the midpoint of the corresponding quantile bin (q_{k-1}, q_k) , for (a, b) chosen to span the 5% of midparents centered on: (red) the 10% quantile, (black) the median, and (green) the 90% quantile of midparent value.

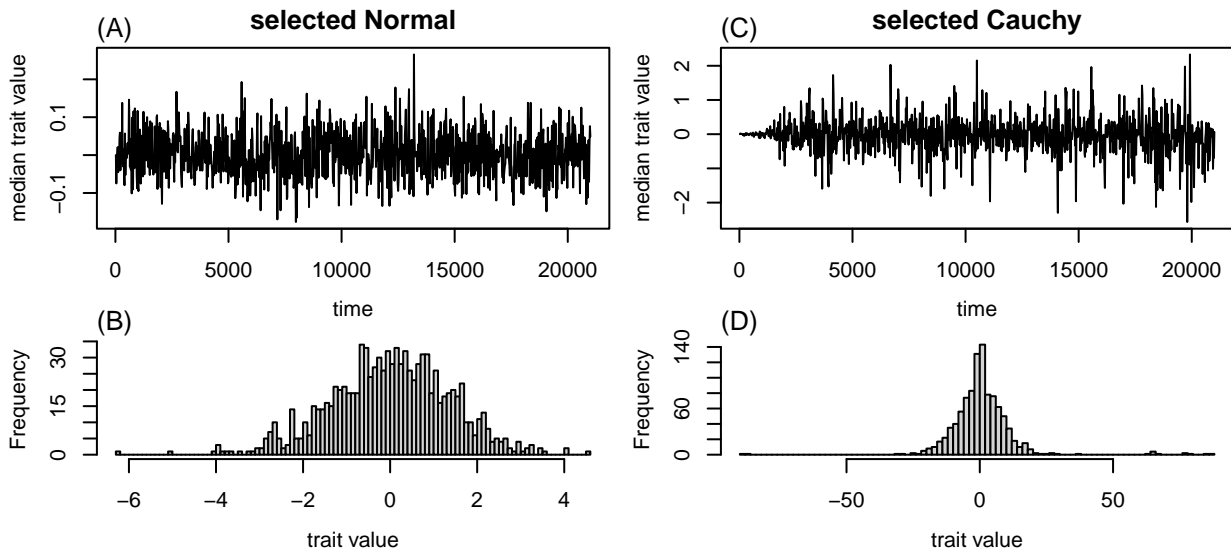


Figure 5: As in Figure 3, except under a model of stabilizing selection (see text for details).

we have seen that the assumption of independence in the infinitesimal model implies a sort of “blending inheritance” for the large effect alleles, and so such work must be accompanied by simulations to see if the predictions are borne out under concrete models of genetic inheritance. Another open question is whether there is some set of realistic assumptions (perhaps involving epigenetic effects, pleiotropy, and/or environmental interactions with genetic effects) which leads to these non-Gaussian infinitesimal models. Finally, it may well be that there is a clever way to set up a trait-only model of evolution that is both insensitive to the underlying genetics and does not rely on the Gaussian central limit theorem.

Finally, we note that above we’ve most often used the Cauchy distribution not because it seems most realistic (indeed data suggests the Lévy stable distribution with $\alpha = 3/2$ might be better) but for mathematical convenience.

Acknowledgements

Thanks go to Nate Pope for useful discussion and to Gregor Gorjanc for insight into modern uses of the animal model and genomic selection. This project was started on a visit by TLP to the University of Oregon that was partially supported by the CNRS PEPS grant “Jeunes Chercheuses et Jeunes Chercheurs”.

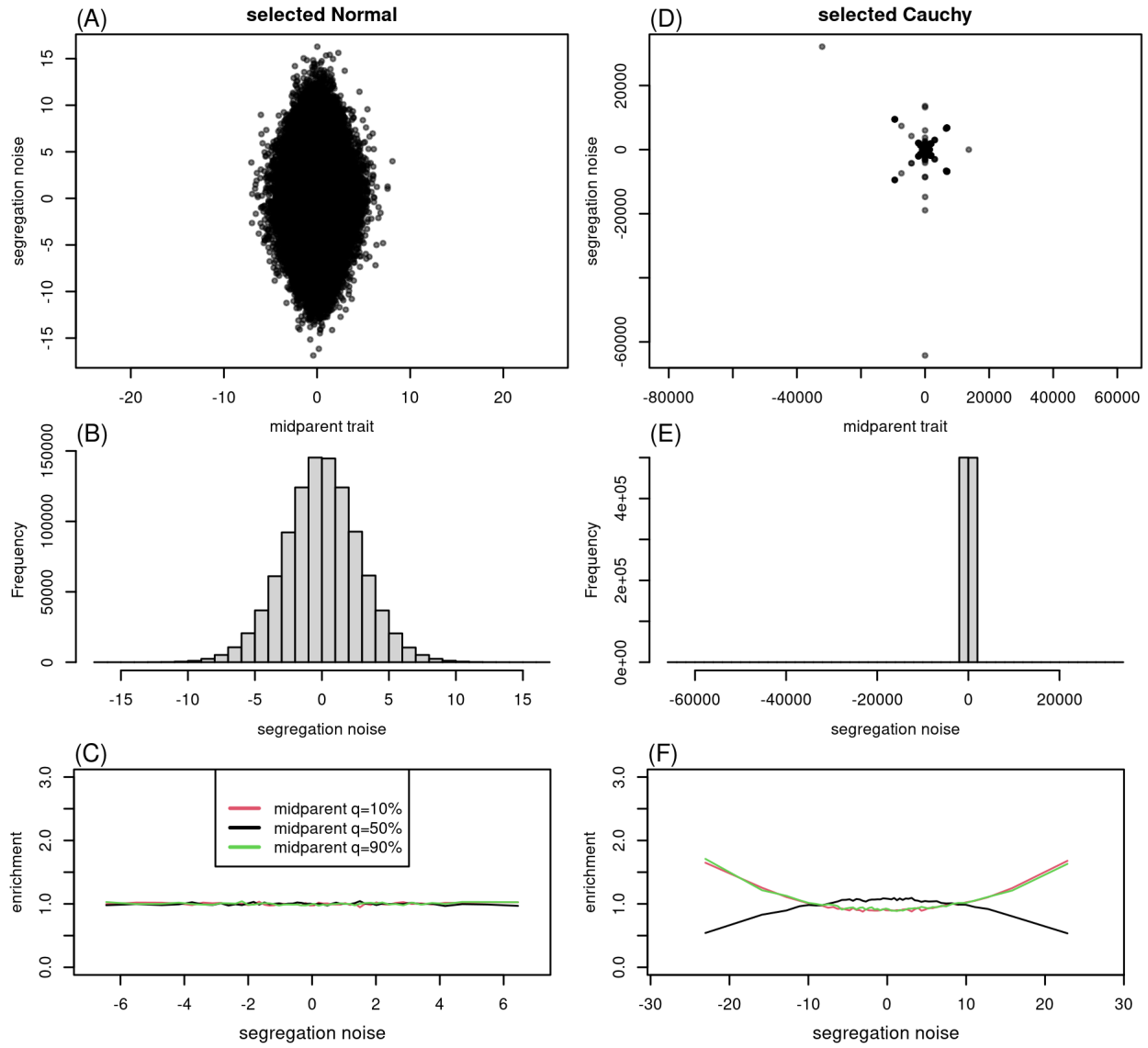


Figure 6: As in Figure 4, except under a model of stabilizing selection (see text for details).

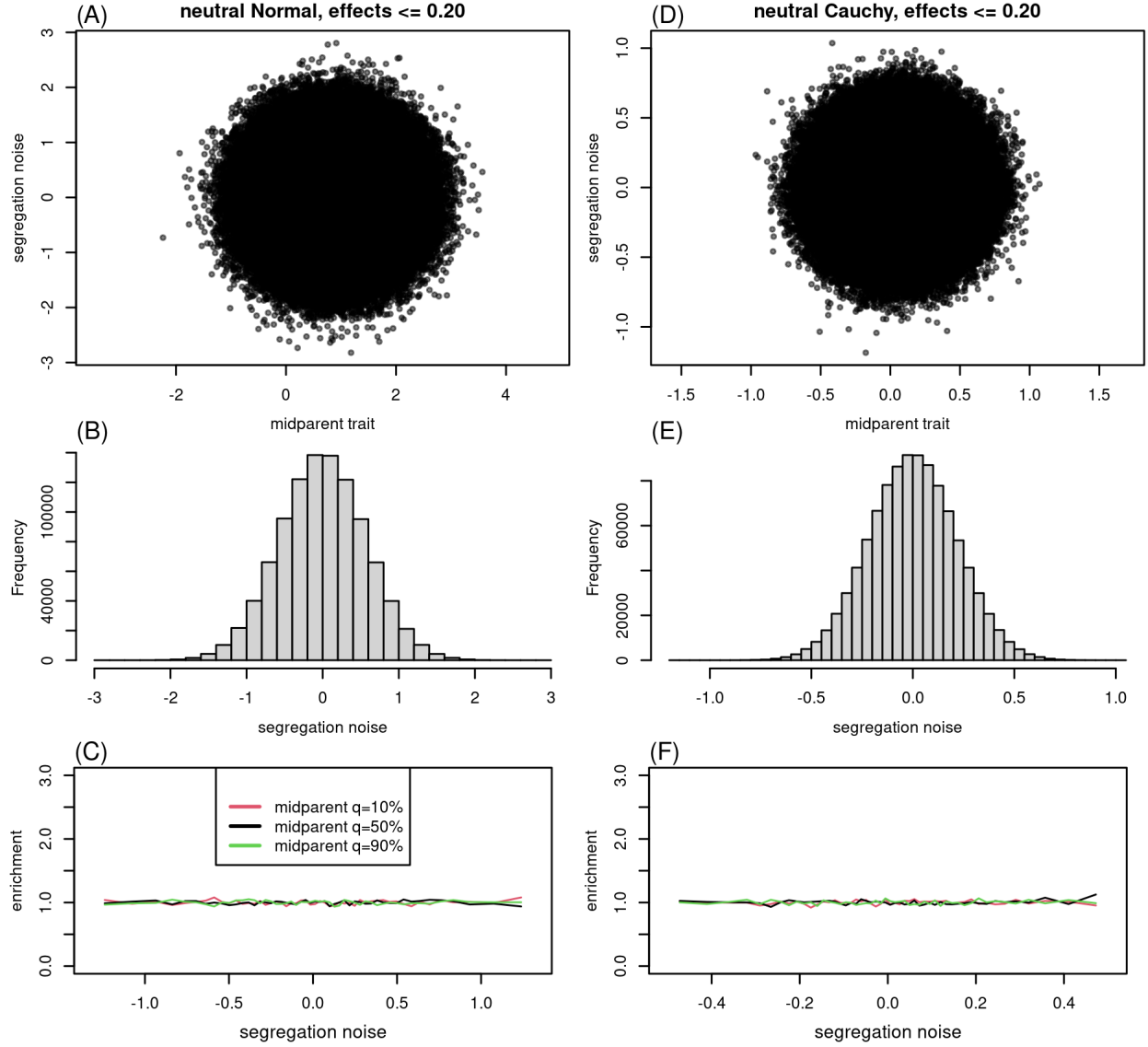


Figure 7: As in Figure 4, except computed using only mutations with absolute value effect size less than 0.2. The same quantities for the simulations of Figure 6 with stabilizing selection are shown in Figure S2.

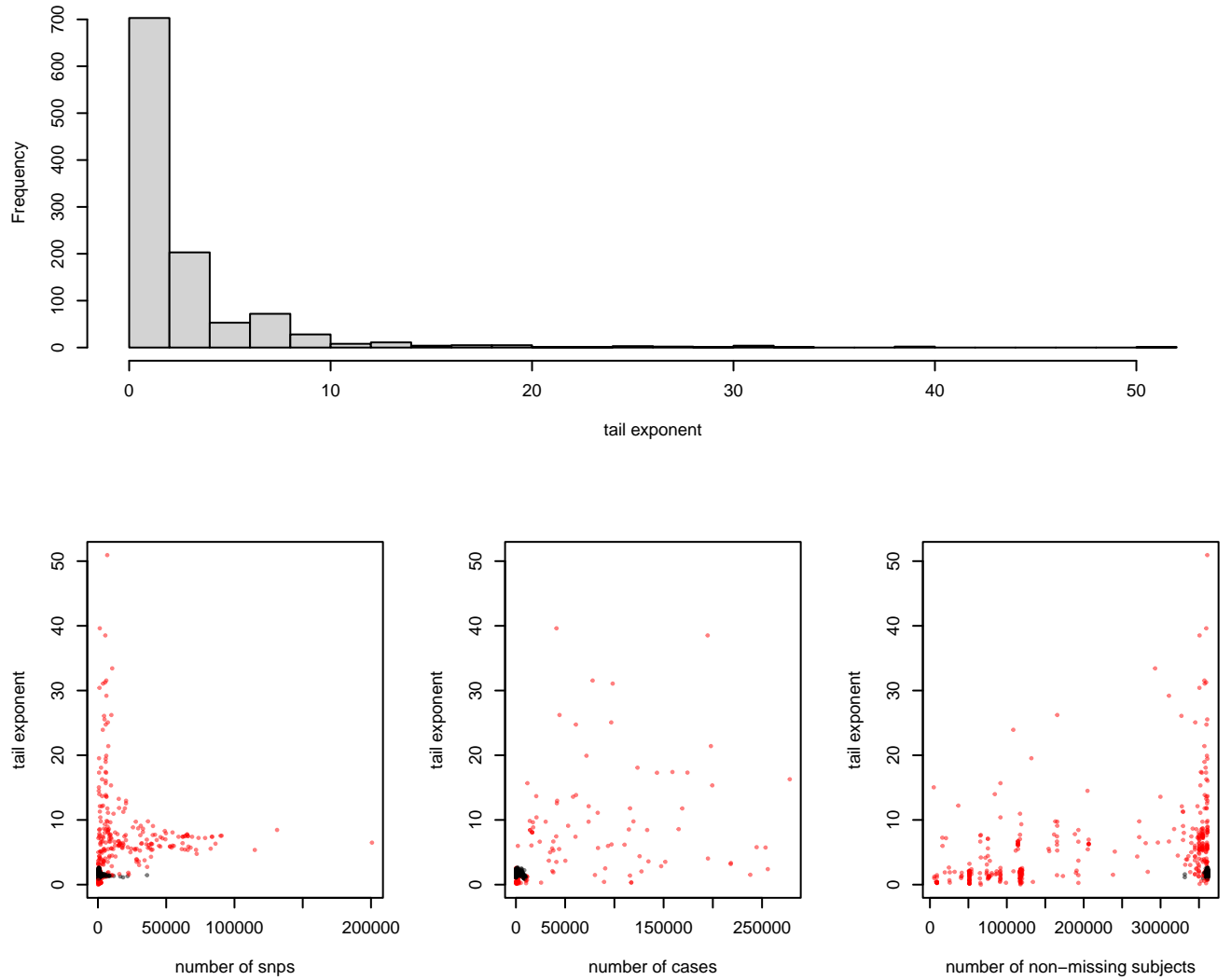


Figure S1: Estimated values of the tail exponent, α , for 1108 illness-related binary phenotypes, including the 333 phenotypes removed in filtering, which are shown as red points in lower plots. (A) distribution of values; and plotted against (B) number of SNPs, (C) number of cases, and (C) number of non-missing subjects.

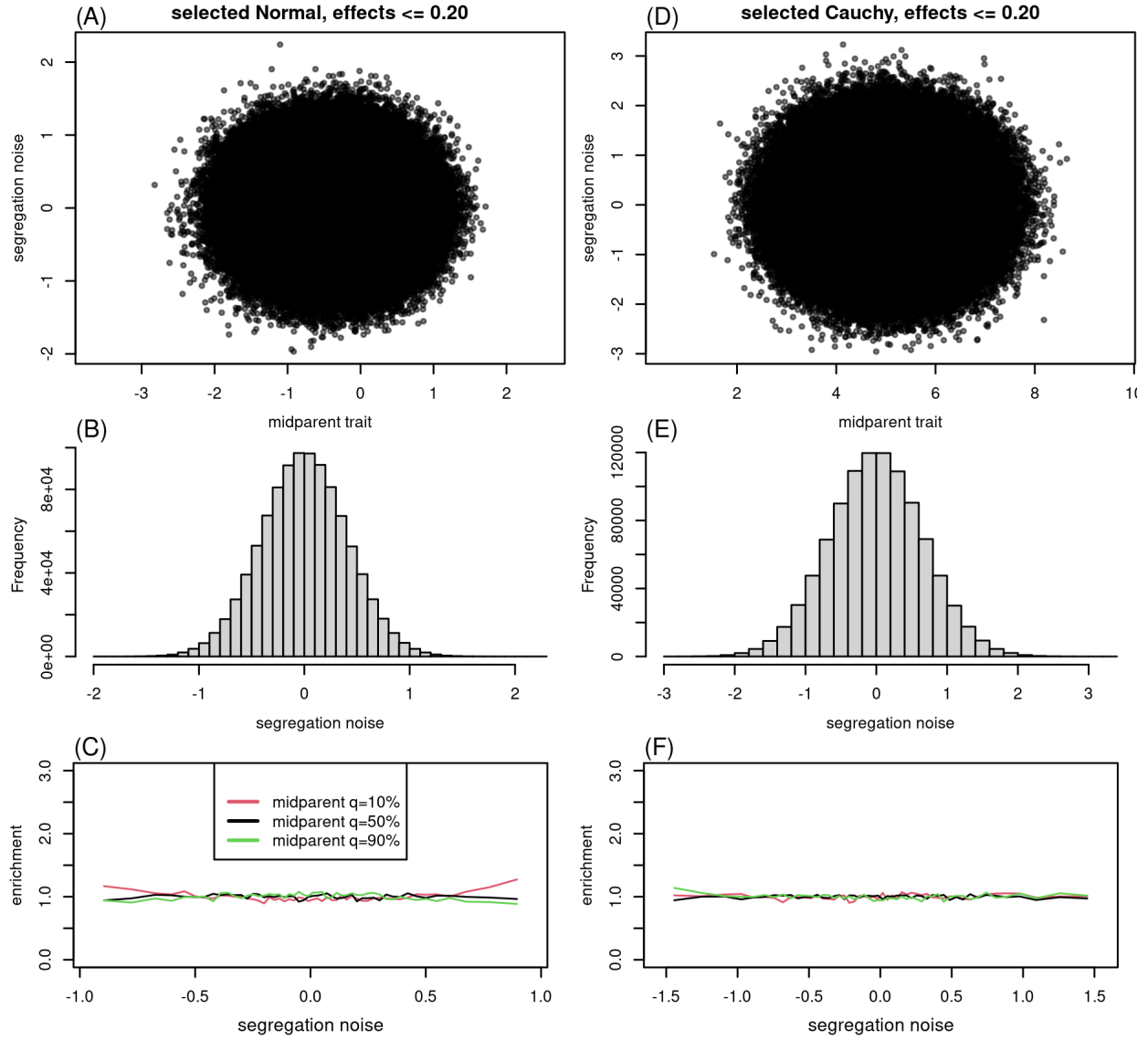


Figure S2: As in Figure 4, except computed using only mutations with absolute value effect size less than 0.2.