

The infinitesimal model of phenotypic evolution:

a microscopic approach

with N. Barton & A. Etheridge

Toulouse, 08/10/2017

The infinitesimal model is a famous model in quantitative genetics. It was introduced by Fisher in 1918. Essentially, it says that if a trait is the result of the very small contribution of many genes, then the distribution of the trait of the offspring of a pair of individuals is normally distributed, with mean the average trait of the parent and with a variance independent of the parental traits. But many versions (all slightly different) of this model exist and to my knowledge, there was no clear statements of the conditions under which this normal distribution is a good approximation for the trait distribution within a family or in the whole population. Our work tries to clarify these conditions, not exhaustively.

So, suppose we're interested in a trait which can vary continuously and which is made of the small contribution of many genetic factors (or "loci" later). Ex: height, life expectancy, ...
→ hence the name "infinitesimal model".

To simplify the question, suppose that these effects are purely additive, in the sense that

$$\underset{\substack{\uparrow \\ \text{trait}}}{Z^{\text{obs}}} = \overbrace{\sum_{i \sim \infty} A_i}^{Z \text{ (genetic component)}} + \underset{\substack{\uparrow \\ \text{"environmental" component}}}{E}$$

$\uparrow$
i-th allelic effect

Here no interactions between loci (no epistatic effects) and we'll also suppose that there are no interactions between the 2 alleles at a given locus (no dominance). We can add them under some conditions on these contributions. Here we remain in a haploid world. ^{related to genes}
but 2 parents

We'll give the details of the model later, but if we assume that an offspring inherits about half of the A_i 's from its father and the other half from its mother, and that the A_i have a variance of order $\frac{1}{M}$ (where M is the number of loci), then we expect that the genetic part of the offspring's trait should indeed be Gaussian, centered at the average of the parental trait (the variance is less clear). The global trait distribution then depends on the form of the environmental contribution E , which we'll suppose to be Gaussian and independent of the rest.

$$\Rightarrow E \sim \mathcal{N}(0, \sigma_E^2)$$

trying to understand the conditions under which the model holds
Most studies have focused on the trait distribution within the whole population (assuming a very large population, usually). In particular, it is used in practical animal breeding to see the effects of ^{natural or artificial} selection on the trait distribution/diversity in the population and so it is important to understand

whether ~~natural~~ selection distorts or not this Gaussian distribution. We'll show that the trait distribution within the whole pop may not be Gaussian due to selection or structure (but the within family distribution (and on the parental traits) is Gaussian)

Our model, finally (can be extended in many ways) Finite population
M loci
no mutation

A_i^m, A_i^f i-allelic effect of the parent

$A_i = A_i^m$ with prob $\frac{1}{2}$ independently between the
 A_i^f $\frac{1}{2}$ loci

In the "base" population,

In the ^{reference} base population, for each i the allelic effect at locus i in each of the N_0 individuals is a i.i.d. draw from some

distribution of the form $a_i(\omega) + \frac{1}{\sqrt{M}} \hat{\eta}_i$ where the $\hat{\eta}_i$ have mean 0
 $\uparrow$
loci

We also suppose that there exists a constant B bounding the support of all $\hat{\eta}_i$'s.

Finally, $E \sim \mathcal{N}(0, \sigma_E^2)$
↑
environmental variance

Combining all this: for a given individual j in generation t , we have

$$\begin{aligned} Z^{obs,j} &= \underbrace{\bar{z}_0}_{\sum_{i=1}^M z_i(0)} + \frac{1}{\sqrt{M}} \sum_{i=1}^M \eta_i^j + E^j \\ &= \bar{z}_0 + \frac{1}{\sqrt{M}} \sum_{i=1}^M \left(X_i^j \eta_i^{mg} + (1-X_i^j) \eta_i^{fg} \right) + E^j \quad \text{(Mendelian inheritance)} \end{aligned}$$

where the X_i^j are i.i.d. in i, j , Bernoulli ($\frac{1}{2}$)

$$= Z^j + E^j$$

Result 1: Knowing the pedigree (ie, the ancestral relationships) of the population up to generation t , together with all the ancestral observed traits up to generation $t-1$, then in the limit as $M \rightarrow \infty$ the genic component of the trait of the offspring of a given pair of individuals in generation $t-1$ are i.i.d. and normally distributed, with mean $\frac{Z^f + Z^m}{2}$ (the average of the genic components of the parental traits, and variance $\sigma_A^2 (1 - F_{m,f}(t-1))$, where $\sigma_A^2 = \lim_{M \rightarrow \infty} \frac{1}{2M} \sum_{i=1}^M \text{Var}(\eta_i)$ and $F_{m,f}(t-1)$ is the probability that the two parents are identical by descent at a given locus.

The law of the observed trait of the offspring can then be computed, especially if the environmental noise is taken to be Gaussian.

In addition, for the j -th individual in generation t , one has for M large enough

$$\left| P\left(\frac{Z_{gen}^i - \frac{Z_{gen}^{p(i)} + Z_{gen}^{m(i)}}{2}}{\sqrt{\frac{\Sigma_t}{M}(1-F_{m,j}(t-1))}} \leq y \mid \begin{array}{c} \text{pedigree} \\ \text{ancestral} \\ \text{traits} \end{array} \right) - P(\mathcal{N}(0,1) \leq y) \right| \leq \frac{t}{\sqrt{M}} C(\Sigma_t, \Delta_t),$$

where Σ_t is the minimum within-family variance ^{up to generation t} and Δ_t is the maximum of the differences $\left| Z_{gen}^i - \frac{Z_{gen}^f + Z_{gen}^m}{2} \right|$ up to generation $t-1$.

By these results, we have found some rather general conditions under which the infinitesimal model holds and we also have a quantitative idea of the speed at which this convergence happens as $M \rightarrow \infty$. This depends on not seeing too extreme traits (Δ_t) and not having too much inbreeding (Σ_t).

(local CLT by Rinoth (1994))

The proof relies on showing that the allele distribution at a given locus is only very slightly modified by conditioning on the observed trait, so that knowing the trait of the ancestors doesn't change much the distribution $\hat{X}_i$ inherited from the ancestors in the base population. Now, some variability ^{at a locus} is lost within a family when the 2 parents

have the same allele inherited from a common ancestor, hence the $(1-F_{m,j})$ ^{but also space for instance}

Finally, we can conclude on the impact of selection ^{but also space for instance}. Knowing the ancestral traits and the pedigree, the traits in generation t are indeed normally distributed with a variance independent ^{within a family} of the parental traits, but only if the way individuals are selected in the past doesn't make Δ_t too large or Σ_t too small.

Then the trait distribution in the whole population may not be Gaussian (it is a mixture of Gaussians), but as long as selection or a spatial structure doesn't distort the pedigree and doesn't produce too extreme traits, then we can use the infinitesimal model to analyse the evolution of the pop.