The infinitesimal model for polyploid and mixed-ploidy populations

Arthur Zwaenepoel

Contents

The basic infinitesimal model

The infinitesimal model assumes that offspring trait values z_{ij} of a parental pair with trait values z_i and z_j are normally distributed:

$$z_{ij} \sim \mathcal{N}\left(\frac{z_i + z_j}{2}, V_{ij}\right)$$

where V_{ij} is called the *segregation variance* and is determined by the hereditary process. The basic infinitesimal model can be derived as the limit of a model where

a quantitative trait is controlled by a large number n of Mendelian loci with additive gene action, each of small effect $\sim O(\sqrt{n})$.

Importantly, V_{ij} is not a function of z_i or z_j, but can evolve over time. If populations are finite, inbreeding will cause the segregation variance to decrease as a function of the inbreeding coefficients of the parental individuals.

A slightly different, and perhaps more insightful, way to specify the same model is to write $z_{ij} = X_i + X_j$, where X_i is the contribution of parent i to the genotypic value of the offspring and X_j the same for j. That is, X_i is the genotypic value of a gamete from i. For gametes produced by a normal meiotic division, we assume $X_i \sim \mathcal{N}(z_i/2, V_i)$, where V_i is the contribution to the segregation variance from i. We therefore have $V_{ij} = (V_i + V_j)/2$. This way of formulating the model stresses that segregation occurs independently in both parents, contributing additively to the segregation variance (which is the variance among offspring within a family).

While the basic phenotypic model holds for arbitrary ploidy levels, the evolution of the segregation variance in finite populations differs for different ploidy levels. In addition, when considering mixed-ploidy populations, we need to consider how equilibrium variances in the model scale with ploidy level, and how offspring of different ploidy levels are derived from a parental pair.

Haploids and diploids

For haploids the recursion for the IBD coefficients in terms of the pedigree matrix is

$$F'_{i,j} = \begin{cases} \sum_k \sum_l P_{i,k} P_{j,l} F_{k,l} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

For diploids, it should be:

$$F'_{i,j} = \sum_k \sum_l P_{i,k} P_{j,l} \begin{cases} F_{k,l} & \text{if } k \neq l \\ \frac{1}{2}(1 + F_{k,k}) & \text{if } k = l \end{cases}$$

Although I wonder whether this is correct for $F'_{i,i}$, since this is defined as the probability of IBD of distinct homologs in i. Consider an individual i offspring

from k and l, then there will be a term $P_{i,k}P_{i,k}\frac{1}{2}(1+F_{k,k}) = \frac{1}{8}(1+F_{k,k})$ in the sum for $F_{i,i}$, which makes no sense?

Tetraploids without double reduction

The segregation variance

Under inbreeding, the expected segregation variance for a family with parents i and j of identical ploidy levels can be decomposed into a contribution from both parents

$$\mathbb{E}[V_{ij}] = \mathbb{E}[V_i] + \mathbb{E}[V_j]$$

We will determine $\mathbb{E}[V_i]$. We define V_0 to be the segregation variance in the base population consisting of unrelated individuals, so that $\mathbb{E}[V_i] = V_0/2$ in the absence of inbreeding.

Assume the parents of i in the previous generation were k and l and consider the contribution of a single locus to $\mathbb{E}[V_i]$. We can consider three mutually exclusive patterns of ancestry:

1. i transmits both genes it inherited from k (w.p. $1/6$)
2. i transmits both genes it inherited from l (w.p. $1/6$)
3. i transmits two genes inherited from distinct parents (w.p. $2/3$)

When the two genes transmitted are IBD, they do not contribute to the segregation variance. Therefore, the expected contribution to the segregation variance for a single locus $\mathbb{E}[v_i]$ conditional on, for instance, scenario 1 above would be

$$\mathbb{E}[v_i] = 0 \times F_{k,k}(t-1) + (v_0/2)(1 - F_{k,k}(t-1))$$

Where $v_0 = V_0/n$ is the average per-locus segregation variance in the base population. Combining all cases, we get

$$\begin{aligned}\mathbb{E}[V_i] &= \frac{1}{6}\frac{V_0}{2}\big(1 - F_{k,k}(t-1)\big) + \frac{1}{6}\frac{V_0}{2}\big(1 - F_{l,l}(t-1)\big) + \frac{2}{3}\frac{V_0}{2}\big(1 - F_{k,l}(t-1)\big) \\ &= \frac{V_0}{2}\left(1 - \frac{1}{6}\big(F_{k,k}(t-1) + F_{l,l}(t-1) + 4F_{k,l}(t-1)\big)\right) \\ &\equiv \frac{V_0}{2}G_i\end{aligned}$$

Note that G_i as defined here is the probability that the two genes transmitted by i to its offspring are not IBD. If i had parents k and j:

$$G_i = 1 - \frac{1}{6}\big(F_{k,k} + F_{l,l} + 4F_{k,l}\big)$$

Where the inbreeding coefficients are from the generation of the parents of i. The total segregation variance for the parental pair (i, j) will be

$$\mathbb{E}[V_{ij}] = V_0\left(\frac{G_i + G_j}{2}\right)$$

That is, in autotetraploids that do not undergo double reduction, the infinitesimal model (as a limit of Mendelian additive loci) under inbreeding is defined as the model where offspring trait values of parents i and j follows a Gaussian distribution:

$$z_{ij} \sim \mathcal{N}\left(\frac{z_i + z_j}{2}, V_0\left(\frac{G_i + G_j}{2}\right)\right)$$

Where V_0 is the segregation variance in the base population and where G_i is the probability that the two genes at a locus transmitted by i are not IBD. Note again that G_i is a function of the inbreeding coefficients of the *parents* of i, so that in the tetraploid model, the inbreeding coefficients of *grandparents* are needed for computing the segregation variance.

Recursions for the inbreeding coefficients

$$F_{i,i}(t) = \sum_k \sum_l P_{i,k} P_{i,l} \frac{1}{6} \big(F_{k,k}(t-1) + F_{l,l}(t-1) + 4F_{k,l}(t-1) \big)$$
$$F_{i,j}(t) = \sum_k \sum_l P_{i,k} P_{j,l} F^*_{k,l}(t-1)$$

where

$$F^*_{k,l}(t) = \begin{cases} F_{k,l}(t) & \text{if } k \neq l \\ \frac{1}{4}\big(1 + 3F_{k,k}(t)\big) & \text{if } k = l \end{cases}$$

To see the latter, note that $k = l$ means that i and j share a parent, in which case for a given gene in i, there is a $1/2$ chance it came from te shared parent, in which case there is a $1/2$ chance that j inherited the same gene independently and there is a $1/4$ chance that this gene is picked in j, so that the probability a given gene is IBD with a random gene in j is $1/16 = P_{i,k}P_{j,k}(1/4)$. When the genes trace back to distinct lineages in the same parent, with probability $P_{i,k}P_{j,k}(3/4)$, they are IBD with probability $F_{k,k}(t-1)$. Note that the probability that both genes trace back to the same parent is embodied in the pedigree matrix P, while $F^*_{k,k}$ is the probability of being IBD *conditional* on tracing to the same parent.

The recursions for the inbreeding coefficients and the law for the reduction of segregation variance with inbreeding together enable simulation of the infinitesimal model for tetraploid populations (without double reduction). I verified the correctness of the derived recursions by comparing simulations from the infinitesimal model to individual-based simulations of a random mating tetraploid population with a large number ($n = 1000$) of unlinked additive Mendelian loci of small effect ($\sim O(1/\sqrt{4n})$.

Double reduction

When tetravalents are formed, a form of internal inbreeding may occur as a result of the phenomenan called double reduction. Schematically, double reduction for a genotype $ABCD$ could look like:

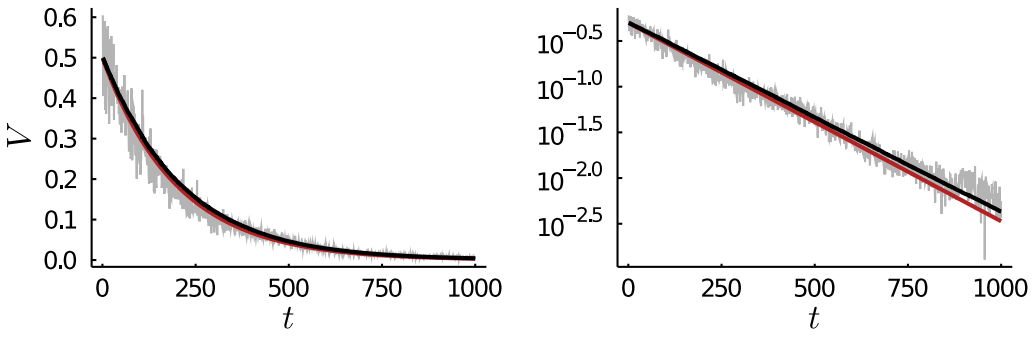


Figure 1: Comparison of the evolution of the segregation variance under inbreeding for the infinitesimal model (black line) and the 1000 unlinked additive loci model (grey). A population of 50 individuals is simulated over 1000 generations. In red the approximation $V_0 e^{-t/4N}$ is shown.

<a scheme>

I don't think this is right Assume double reduction occurs at any locus with probability d (what *exactly* happens with probability d?). When a single double reduction takes place at a locus, helf of the gametes will contain a pair of IBD genes at the locus, so that the segregation variance becomes $\frac{v_0}{2}$. Following this reasoning, the segregation variance can be found as

$$\mathbb{E}[V_i] = (1-d)^2\mathbb{E}[V_i^*] + 2d(1-d)\frac{\mathbb{E}[V_i^*]}{2}$$

where $\mathbb{E}[V_i^*]$ is the contribution to the segregation variance of individual i in the absence of double reduction. This verifies that double reduction leads to a kind of 'internal' inbreeding, as the segregation variance is reduced with respect to V_i^*.

The infinitesimal model for mixed-ploidy populations

Scaling of genetic variance across ploidy levels

If we consider the infinitesimal model as the limit of a large number of additive Mendelian loci, the genotypic value of an individual is defined as

$$z = \sum_{k=1}^{n} \sum_{l=1}^{m} a_{k,l}$$

where $a_{k,m}$ is the allelic effect of homolog l of locus k. Clearly, if we assume identical allelic effects in diploids and congeneric polyploids, the phenotypic range of tetraploids will be double that of diploids. While the assumption of additive allelic effects is of course in itself problematic, conditional on this assumption, the assumption of *equal* additive effects across ploidy levels seems rather unbiological.

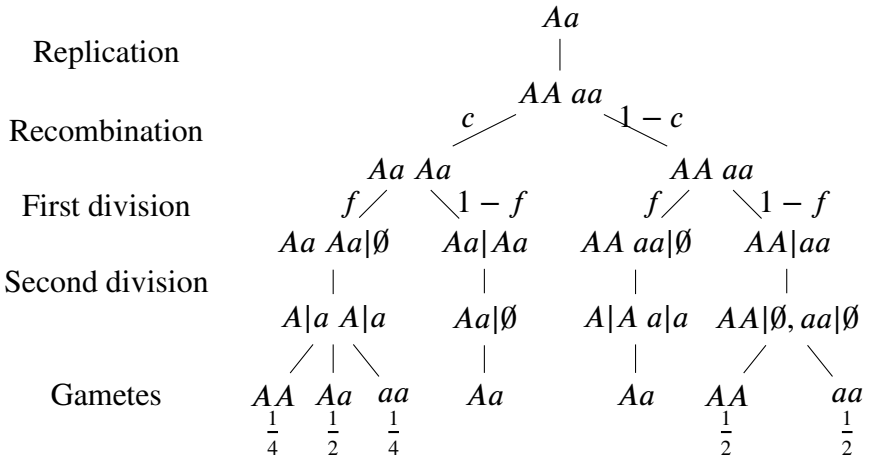
We introduce a scaler for allelic effects in an m-ploid $\sqrt{\beta_m}$, so that

$$z = \sqrt{\beta_m} \sum_{k=1}^{n} \sum_{l=1}^{m} a_{k,l}$$

where we define $\beta_2 = 1$, so that $a_{k.}$ is the effect the allele would have in a diploid individual. With this model the HWLE additive genetic variance $V_{A,m}$ and the base segregation variance $V_{0,m}$ in an m-ploid population are $\frac{m}{2}\beta_m V_{A,2}$ and $\frac{m}{2}\beta_m V_{0,2}$ respectively.

Unreduced gamete formation in diploids

Naively, one may think that an unreduced gamete contains the parental genome, and that as a result the segregation variance for a $2n \times 2n \rightarrow 4n$ cross would be 0. However, the mechanisms of unreduced gamete formation do not necessarily lead to a faithful transmission of the complete diploid genome. Unreduced gametes are formed in two ways, depending on the meiotic abberation that leads to their origin: (1) first division restitution (FDR) of (2) second division restitution (SDR). Consider a locus in a diploid with two distinct genes A and a. Assume recombination happens with probability c and that conditional on unreduced gamete formation, formation is due to FDR with probability f while it is due to SDR with probability $1 - f$. The different unreduced gametes that are formed are represented in the following diagram:



Clearly, the mechanism of unreduced gamete formation creates segregation variance, as not all random tetraploid offspring from a single diploid parental pair will receive the same pair of genes from each parent, depending on whether or not recombination has occurred and FDR rather than SDR generates the unreduced gamete.

If we consider the diploid genotype Aa depicted in the diagram above and consider X_A the number of A alleles transmitted, we can easily find

$$\operatorname{var} X_A = \left(1 - c - f + \frac{3}{2}cf\right) := \xi$$

which is also the probability of transmitting two copies of the same allele. We shall denote this derived quantity as ξ.

Of course, if the locus is IBD with probability $F_{i,i}$, we have that the expected value of $\operatorname{var} X_A$ is $\xi(1 - F_{i,i})$. Under the model delineated above with the scaling of allelic effects across ploidy levels, we find that the contribution of a single parent i to the segregation variance among tetraploid offspring of a diploid parental pair is

$$V_i = 2\beta_4 V_{0,2}\xi(1 - F_{i,i})$$

Notably, the mean offspring phenotype among tetraploid offspring from a diploid cross can no longer be the average of the two parental phenotypes if we assume the infinitesimal model as the limit of the Mendelian unlinked additive loci model. The expected contribution of a single parent i will be $X_i = \sqrt{\beta_4} z_i$ so that for a cross of two diploids via unreduced gametes

$$z_{ij} \sim \mathcal{N}\left(\sqrt{\beta_4}(z_i + z_j), V_i + V_j\right)$$

Inbreeding coefficients in the mixed-ploidy system

In a mixed-ploidy system tracking inbreeding coefficients becomes slightly more complicated, as our recursions will differ whether some individual is derived from parents of the same cytotype or not.

Recall that ξ is the probability that an unreduced gamete transmits two copies of the same allele at some locus. We still have that for a tetraploid individual i

$$F'_{i,i} = \sum_k \sum_l P_{i,k} P_{i,l} \frac{1}{6}\left(F^*_{k,k} + F^*_{l,l} + 4F_{k,l}\right)$$

where

$$F^*_{k,k} = \begin{cases} F_{k,k} & \text{if } m_k = 4 \\ F_{k,k} + \xi & \text{if } m_k = 2 \end{cases}$$

For $F'_{i,j}$ we still have

$$F'_{i,j} = \sum_k \sum_l P_{i,k} P_{i,l} F^*_{k,l}$$

where now

$$F^*_{k,l} = \begin{cases} F_{k,l} & \text{if } k \neq l \\ \frac{1}{2}(1 + F_{k,k}) & \text{if } k = l, m_k = 2 \\ \frac{1}{4}(1 + 3F_{k,k}) & \text{if } k = l, m_k = 4 \end{cases}$$

(I think)

Lastly we need G_i, the probability that the two genes transmitted by a tetraploid individual i to an offspring are not IBD.

$$G_i = 1 - \frac{1}{6}\left(F^*_{k,k} + F^*_{l,l} + 4F_{k,l} \right)$$

where

$$F_{k,k} = \begin{cases} F_{k,k}(1-\xi) + \xi & \text{if } m_k = 2 \\ F_{k,k} & \text{if } m_k = 4 \end{cases}$$

I think that's it (ignoring triploids and double reduction for now). Not sure how to organize the calculations yet though.