The infinitesimal model for polyploid and mixed-ploidy populations

Arthur Zwaenepoel

Contents

The basic infinitesimal model

Phenotypic model definition

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

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

where $V_{i,j}$ 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 $\sim O(\sqrt{n})$.

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. Segregation occurs independently in both parents, contributing additively to the segregation variance $V_{i,j} = V_i + V_j$.

Inbreeding

Importantly, $V_{i,j}$ is not a function of z_i or z_j (although they may be correlated), but can evolve over time. If populations are finite, inbreeding will lead to drift and cause the segregation variance to decrease as a function of the relatedness of the parental individuals. Importantly, while the basic phenotypic model holds for arbitrary ploidy levels, genetic drift – and consequently, the evolution of the segregation variance in finite populations – will differ for different ploidy levels.

Haploids and diploids

The infinitesimal model for finite populations of haploid and diploid individuals is described in detail in @barton2017. Let $F_{i,j}$ be the inbreeding coefficient for a pair of individuals (i, j), defined as the probability that a gene in i is identical by descent (IBD) to a gene in j. We define $F_{i,i}$ to be the probability that two *distinct* genes in i are IBD. In the haploid case, the segregation variance for a parental pair (i, j) is reduced with inbreeding to $V_{ij} = V_0(1 - F_{i,j})$, where V_0 is the segregation variance in the base population consisting of unrelated individuals. For a diploid pair, the segregation variance is

$$V_{ij} = V_i + V_j = \frac{V_0}{2}(1 - F_{i,i}) + \frac{V_0}{2}(1 - F_{j,j}) = V_0\left(1 - \frac{F_{i,i} + F_{j,j}}{2}\right)$$

When simulating the infinitesimal model, we need a way to efficiently track the inbreeding coefficients during the simulation. To do so, we will make use of the pedigree matrix P as in @barton2017. For haploids the recursion for the inbreeding coefficients in the offspring generation (F') in terms of P and the inbreeding coefficients in the preceding generation (F) is

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

As given by @barton2017. The same authors give for diploids

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

But this is incorrect[1] for $i = j$, since $F_{i,i}$ is defined as the probability of identity by descent (IBD) of *distinct* homologs in i. The correct expression for the diagonal elements of F is

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

where k and l are the parents of i.

Tetraploids without double reduction

We define $F_{i,i}$ in a tetraploid individual i as the probability that two randomly picked distinct genes at some locus in i are IBD. With this definition we can easily see that, in the absence of double reduction, the segregation variance contributed by i is

$$V_i = \frac{V_0}{2}(1 - F_{i,i})$$

[1] Consider an individual i which is an offspring from k and l, with $k \neq l$. There will be a term $P_{i,k}^2 \frac{1}{2}(1 + F_{k,k}) = \frac{1}{8}(1 + F_{k,k})$ as well as a term $P_{i,l}^2(1 + F_{l,l})$ in the sum for $F_{i,i}$, both of which are spurious since $F_{i,i}$ is the probability that two distinct genes are IBD, and the probability that two *distinct* genes come from parent k is not $P_{i,k}^2$ but 0.

If i had parents k and l, we can, given the F values of the parental generation, compute $F'_{i,i}$. To do so, we consider three mutually exclusive patterns of ancestry: two randomly picked distinct genes in i are either (1) both inherited from k (w.p. $1/6$), (2) both inherited from l (w.p. $1/6$) or (3) each inherited from a different parent (w.p. $2/3$). Clearly, if $k \neq l$ and when there is no double reduction:

$$F'_{i,i} = \frac{1}{6}F_{k,k} + \frac{1}{6}F_{l,l} + \frac{3}{2}F_{k,l}$$

If $k = l$ we find that $F'_{i,i} = \frac{1}{4}(1 + 3F_{k,k})$. For all other (non-diagonal) entries in F, a recursion similar to the diploid case is found

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

with

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

It turns out that for a population of m-ploids, where $m \in \{1, 2, 4\}$, the update rule for F can be written succintly in matrix notation.

$$F' = P\bigg(F + \frac{1}{m}\big(I - \text{diag}F\big)\bigg)P^T$$

but with the diagonal elements given by $F'_{i,i}$.

Double reduction

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

<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^*.

Simulations

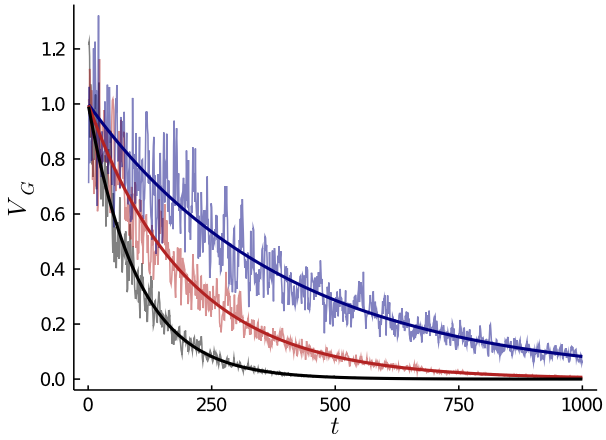


Figure 1: Decline of genetic variance (V_G) with inbreeding (drift) in a population of 100 haploids (black), 100 diploids (red) and 100 tetraploids (blue) with equal V_0 and initial V_G. Solid lines show the (approximate for tetraploids) expected exponential decline $V_G(0)e^{-t/(mN)}$.

The infinitesimal model for mixed-ploidy populations

Scaling of genetic variance across ploidy levels

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. 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

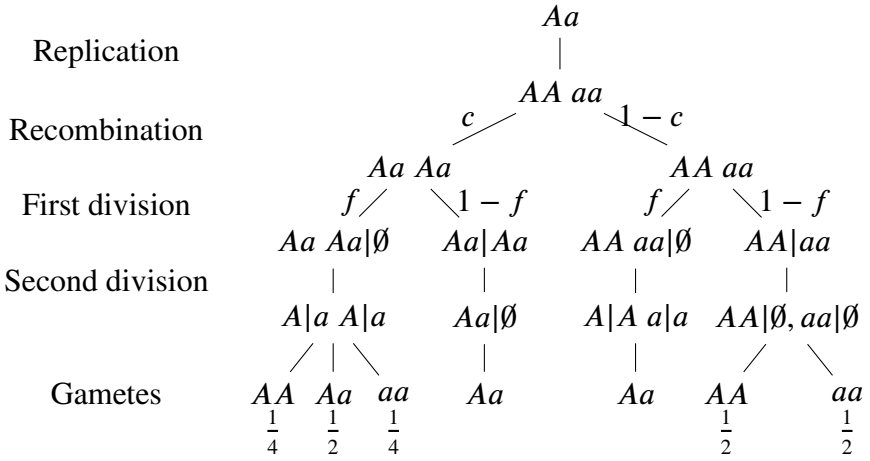
$$z = \sqrt{\beta_m} \sum_{k=1}^{n} \sum_{l=1}^{m} a_{k,l}$$

where we define $\beta_2 = 1$, so that $a_{k\cdot}$ 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 zero. 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 diploid parent i to the segregation variance among tetraploid offspring 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. Let m_k denote the ploidy level of individual k. We still have that for a tetraploid individual i

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

where now

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

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), so that we still have a matrix expression

$$F' = P\Big(F + \big(I - m\text{diag}F\big)\Big)P^T$$

still holds, but where m is now a vector where the kth element is $1/m_k$, recording the reciprocal of the ploidy levels in the parental generation.