

1 Moment generating functions for the distribution of quantitative trait values

1.1 A generalized form

1.1.1 Stepping stone mutation model

A basic question in quantitative genetics is what is the distribution of trait values in a sample of individuals. In their investigation, SCHRAIBER and LANDIS (2015) found a generating function for this distribution under the standard neutral coalescent model and a general stepping stone model of mutation. We wish to extend this framework to additional population models. A first step is to show a connection between the distribution of branch lengths on a genealogy and the distribution of trait values in the sampled individuals. This should be useful because the generating functions under various models of population structure have already been investigated in some depth (LOHSE *et al.*, 2011).

Let \mathbf{T} be a random vector containing the lengths of all possible branches on a coalescent tree. For instance, if we had a three samples a , b , and c , then $\mathbf{T} = \{T_{\{a\}}, T_{\{b\}}, T_{\{c\}}, T_{\{a,b\}}, T_{\{a,c\}}, T_{\{b,c\}}\}$. Let Ω contain all possible configurations that coalescent branches can subtend. For our example, $\Omega = \{\{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}\}$. Let \mathbf{Y} be a random vector containing the quantitative trait values in the sampled individuals. For our example, $\mathbf{Y} = \{Y_a, Y_b, Y_c\}$. It is important to note that these values indicate the change in the quantitative trait since the value in the MRCA of the sample, not their actual values.

The moment generating function of \mathbf{Y} is

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = E[e^{\mathbf{k} \cdot \mathbf{Y}}] = \int e^{\mathbf{k} \cdot \mathbf{Y}} P(\mathbf{Y} = \mathbf{y}) d\mathbf{y}, \quad (1)$$

where \mathbf{k} is a vector of dummy variables for each sample. Although imperfect, this is the notation I will use for writing integrals over probability distributions. Rewriting this expression by conditioning on the genealogy we get

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \int e^{\mathbf{k} \cdot \mathbf{Y}} \int P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) P(\mathbf{T} = \mathbf{t}) d\mathbf{t} d\mathbf{y} \quad (2)$$

$$= \int \int e^{\mathbf{k} \cdot \mathbf{Y}} P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) d\mathbf{y} P(\mathbf{T} = \mathbf{t}) d\mathbf{t} \quad (3)$$

We can write each Y_i as a sum of the change in the trait value occurring on each branch of the genealogy.

$$Y_i = \sum_{\omega \in \Omega} Y_{i,\omega} \quad (4)$$

Of course, many of these branches will not subtend i , so the change along them is defined to be zero. The $Y_{i,\omega}$ are correlated because the underlying branch lengths are correlated, but they are conditionally independent given \mathbf{T} . This is due to the assumption of a stepping stone mutation model and that mutations occur as a poisson process along the branches. We can therefore factor $\int e^{\mathbf{k} \cdot \mathbf{Y}} P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) d\mathbf{y}$ into independent parts. We first write

$$\mathbf{k} \cdot \mathbf{y} = \sum_{\omega \in \Omega} \left(\sum_{i \in \omega} k_i y_{i,\omega} \right) \quad (5)$$

and

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) = \prod_{\omega \in \Omega} P(\mathbf{Y}_\omega = (y_{i,\omega})_{i \in \omega} | \mathbf{T} = \mathbf{t}). \quad (6)$$

This yields

$$\int e^{\mathbf{k} \cdot \mathbf{Y}} P(\mathbf{Y} = \mathbf{y} | \mathbf{T} = \mathbf{t}) d\mathbf{y} = \prod_{\omega \in \Omega} \int \exp \left(\sum_{i \in \omega} k_i y_{i,\omega} \right) P(\mathbf{Y}_\omega = (y_{i,\omega})_{i \in \omega} | \mathbf{T} = \mathbf{t}) d(y_{i,\omega})_{i \in \omega}. \quad (7)$$

To move on we require two results. The first is that the change in a trait value along some branch of length t_ω is a compound Poisson process. The moment generating function for a compound Poisson process over a time t is $\exp(\lambda t(\psi(k) - 1))$. Where λ is the rate that events happen and ψ is the moment generating function for the distribution of mutation effects. The second results is that the moment generating function for the distribution of two completely correlated random variables X_1 and X_2 that have the same distribution is $\varphi_{X_1}(k_1 + k_2)$. Since the trait change along a branch t_ω is the same for all samples subtended by the branch, we can write the product terms in equation 7 as

$$\exp \left(\frac{\theta}{2} t_\omega \left(\psi \left(\sum_{i \in \omega} k_\omega \right) - 1 \right) \right). \quad (8)$$

This gives

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \int \prod_{\omega \in \Omega} \exp \left(\frac{\theta}{2} t_\omega \left(\psi \left(\sum_{a \in \omega} k_a \right) - 1 \right) \right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t}. \quad (9)$$

This is simply the moment generating function for the genealogy \mathbf{T} with $\frac{\theta}{2} (\psi(\sum_{i \in \omega} k_\omega) - 1)$ substituted for the dummy variable of branch T_ω . Or,

$$\varphi_{\mathbf{T}}(\mathbf{s}) \Big|_{s_\omega = \frac{\theta}{2} y(\psi(\sum_{a \in \omega} k_a) - 1)} \quad (10)$$

1.1.2 House of cards mutation model

The model used above is just one possible way in which mutations can affect quantitative trait values. This model assumes that the contribution to the phenotype of an individual from a locus i is a linear combination of the effects of mutations that have occurred at the locus. Moreover, the phenotype of an individual is a linear combination of all the mutations that have occurred in the genome of that individual. The big assumption is that each mutation at a locus adds a random effect to the current effect of that locus. This model is called the stepping stone model and was developed by KIMURA (1965). An alternative model is one in which each mutation causes the effect of that locus to be drawn from some distribution, erasing the effects of all other mutations that had occurred at that locus in the past. This mutational model is called the house of cards model and was developed by KINGMAN (1978).

Here I derive a moment generating function for the house of cards model. This turns out to have a more complicated form than the stepping stone model because the effect of a mutation on an individual's phenotype depends on what other mutations occur at that locus.

As before, the moment generating function is defined as

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = E [e^{\mathbf{k} \cdot \mathbf{Y}}] = \int e^{\mathbf{k} \cdot \mathbf{Y}} P(\mathbf{Y} = \mathbf{y}) d\mathbf{y}. \quad (11)$$

We employ the same tool of conditioning on the genealogy, \mathbf{T} , but we additionally condition on the mutational configuration \mathbf{M} . The mutational configuration is a vector whose entry m_ω is one if a mutation occurs on branch ω and zero otherwise. This gives

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \int e^{\mathbf{k} \cdot \mathbf{Y}} \int \int P(\mathbf{Y} = \mathbf{y} | \mathbf{M} = \mathbf{m}) P(\mathbf{M} = \mathbf{m} | \mathbf{T} = \mathbf{t}) P(\mathbf{T} = \mathbf{t}) d\mathbf{m} d\mathbf{t} d\mathbf{y}. \quad (12)$$

$$= \sum_{\mathbf{m}} \int e^{\mathbf{k} \cdot \mathbf{Y}} P(\mathbf{Y} = \mathbf{y} | \mathbf{M} = \mathbf{m}) d\mathbf{y} \int P(\mathbf{M} = \mathbf{m} | \mathbf{T} = \mathbf{t}) P(\mathbf{T} = \mathbf{t}) d\mathbf{t} \quad (13)$$

Of course, the phenotype values are conditionally independent of the genealogy given the mutational configuration. The first integral is the difficult one because it requires knowing the dependence structure of the phenotypic values given the mutation configuration. Individual phenotypes will be zero if no mutations occur, independent if mutations occur on different branches, and completely correlated if mutations occur on the same branch. Additionally, only the lowest level mutations count. By this we mean for each individual only the mutations on the branch with the smallest number of descendants count. This can be written succinctly as

$$\int e^{\mathbf{k} \cdot \mathbf{Y}} P(\mathbf{Y} = \mathbf{y} | \mathbf{M} = \mathbf{m}) d\mathbf{y} = \prod_{\omega} \psi \left(\sum_{i \in \omega} k_i \mathcal{G}(i, \omega, \mathbf{m}) \right). \quad (14)$$

Where

$$\mathcal{G}(i, \omega, \mathbf{m}) = \begin{cases} 1 & \text{if } \omega \text{ is lowest level branch in } \mathbf{m} \text{ containing a mutation subtending } i \\ 0 & \text{otherwise.} \end{cases}$$

The second integral can be written as

$$\int \prod_{\omega} \left(m_{\omega} + (-1)^{m_{\omega}} e^{-\frac{\theta}{2} t_{\omega}} \right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t}. \quad (15)$$

This multiplies the probability that one or more mutations occur along a branch. We can recognize that this will just be a sum of generating functions for the genealogy evaluated at different $-\frac{\theta}{2} \mathbf{m}$. It can therefore be written as

$$\sum_{\mathbf{m}'} (-1)^{\mathbf{m} \cdot \mathbf{m}'} \mathcal{K}(\mathbf{m}, \mathbf{m}') \varphi_T(-\frac{\theta}{2} \mathbf{m}'). \quad (16)$$

Where

$$\mathcal{K}(\mathbf{m}, \mathbf{m}') = \begin{cases} 1 & \text{if } |\mathbf{m}| + \mathbf{m} \cdot \mathbf{m}' = \text{sum}(\mathbf{m}) + \text{sum}(\mathbf{m}') \\ 0 & \text{otherwise.} \end{cases}$$

This function returns one if \mathbf{m} and \mathbf{m}' contain none of the same zero elements, and zero otherwise.

The moment generating function overall can therefore be written as

$$\sum_{\mathbf{m}} \left(\prod_{\omega} \psi \left(\sum_{i \in \omega} k_i \mathcal{G}(i, \omega, \mathbf{m}) \right) \right) \left(\sum_{\mathbf{m}'} (-1)^{\mathbf{m} \cdot \mathbf{m}'} \mathcal{K}(\mathbf{m}, \mathbf{m}') \varphi_T(-\frac{\theta}{2} \mathbf{m}') \right). \quad (17)$$

1.2 A central limit theorem

Recall that the moment generating function for the distribution of trait values from a single locus is

$$\varphi_{\mathbf{Y}}(\mathbf{k}) = \int \exp \left(\sum_{\omega \in \Omega} s_{\omega} t_{\omega} \right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t} \Big|_{s_{\omega} = \frac{\theta}{2} (\psi(\sum_{a \in \omega} k_a) - 1)}.$$

If we substitute in the Taylor series expansions for the moment generating function of the trait value distribution we get

$$\int \prod_{\Omega} \exp \left[t_{\omega} \frac{\theta}{2} \left(\sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a \right)^n \right) \right] P(\mathbf{T} = \mathbf{t}) d\mathbf{t}.$$

If we then write the Taylor series of each exponential function we get

$$\int \prod_{\Omega} \left[\sum_{j=0}^{\infty} \frac{t_{\omega}^j}{j!} \left(\frac{\theta}{2} \right)^j \left(\sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a \right)^n \right)^j \right] P(\mathbf{T} = \mathbf{t}) d\mathbf{t},$$

which is equivalent to

$$1 + \sum_{\Omega} E[T_{\omega}] \frac{\theta}{2} \sum_{n=1}^{\infty} \frac{m_n}{n!} \left(\sum_{a \in \omega} k_a \right)^n + \sum_{\Omega \times \Omega} \frac{1}{2} E[T_{\omega_1} T_{\omega_2}] \sum_{n=1}^{\infty} \frac{\theta}{2} \frac{m_n}{n!} \left(\sum_{a \in \omega_1} k_a \right)^n \sum_{n=1}^{\infty} \frac{\theta}{2} \frac{m_n}{n!} \left(\sum_{a \in \omega_2} k_a \right)^n + \dots$$

This is raised to the power L for a trait controlled by L loci. We want the limit as the number of loci increases while the size of mutational decreases. This can be expressed by the limits $L \frac{\theta}{2} m_1 \rightarrow \mu$, $L \frac{\theta}{2} m_2 \rightarrow \sigma^2$ and $L \frac{\theta}{2} m_i \rightarrow 0$ for $i > 2$ as $L \rightarrow \infty$. Knowing we will not be retaining m_3 and above we can rewrite the mgf as

$$\left(1 + \sum_{\Omega} E[T_{\omega}] \frac{\theta}{2} \left(m_1 \left(\sum_{a \in \omega} k_a \right) + \frac{m_2}{2} \left(\sum_{a \in \omega} k_a \right)^2 \right) \right)^L.$$

Taking the appropriate limit here is severely annoying to do.

Doing so is tedious, but the result of taking these limits is

$$\exp \left(\sum_{\omega \in \Omega} E[T_{\omega}] \left(\mu \sum_{a \in \omega} k_a + \frac{\sigma^2}{2} \left(\sum_{a \in \omega} k_a \right)^2 \right) \right). \quad (18)$$

This is multivariate normal distribution with mean equal to $E[T_{MRC A}] \mu$, variance equal to $E[T_{MRC A}] \sigma^2$, and covariance between Y_a and Y_b equal to $E[\tau_{a+b}] \sigma^2$. This can be seen from equation (18) by noting that in the mgf for a multivariate normal distribution the coefficient in the exponential of k_a is the mean of Y_a and the coefficient of $k_a k_b$ is $2 \text{Cov}[Y_a, Y_b]$ if $a \neq b$ and $\text{Var}[Y_a]$ if $a = b$. A low mutation rate approximation can be taken by ignoring these.

These Y value can never be directly observed. Rather, differences between individual trait values are what analyses would be based on. Since the Y are normal so will be differences like $Y_a - Y_b$. Conveniently, the second term that depends on the variance of the $T_{MRC A}$ in the variance

and covariance terms of the normal distribution does not appear in the distribution of trait differences.

$$\begin{aligned}\text{Var}[Y_a - Y_b] &= \text{Var}[Y_a] + \text{Var}[Y_b] - 2\text{Cov}[Y_a, Y_b] \\ &= 2(\text{E}[T_{MRC A}]\sigma^2) - 2(\text{E}[\tau_{a+b}]\sigma^2) \\ &= 2(\text{E}[T_{MRC A}] - \text{E}[\tau_{a+b}])\sigma^2.\end{aligned}$$

This also tells us that the distribution of trait differences will be the same in the infinitesimal limit with and without a low mutation rate approximation.

A much simpler heuristic derivation of the limiting normal distribution can be done by calculating the variance and covariance at a single locus. This derivation is very similar to that done by SCHRAIBER and LANDIS (2015). Using the law of total variance we can write

$$\text{Var}[Y] = \text{E}[\text{Var}[Y|T]] + \text{Var}[\text{E}[Y|T]]$$

The variance conditional on T can be calculated again using the law of total variance and conditioning on the number of mutation at the locus.

$$\begin{aligned}\text{Var}[Y|T] &= \text{E}[\text{Var}[Y|M]|T] + \text{Var}[E[Y|M]|T] \\ &= \text{E}[M(m_2 - m_1^2)|T] + \text{Var}[Mm_1|T] \\ &= \frac{\theta}{2}T(m_2 - m_1^2) + \frac{\theta}{2}Tm_1^2 \\ &= \frac{\theta}{2}Tm_2\end{aligned}$$

$$\text{E}[Y|T] = \frac{\theta}{2}Tm_1$$

$$\text{Var}[\frac{\theta}{2}Tm_1] = \left(\frac{\theta}{2}m_1\right)^2 \text{Var}[T]$$

Therefore we have

$$\text{Var}[T] = \frac{\theta}{2}m_2\text{E}[T_{MRC A}] + \left(\frac{\theta}{2}m_1\right)^2\text{Var}[T_{MRC A}]. \quad (19)$$

The same procedure can be done for the covariance.

$$\text{Cov}[Y_a, Y_b] = \text{E}[\text{Cov}[Y_a, Y_b|T]] + \text{Cov}[E[Y_a|T], E[Y_b|T]]$$

We can break Y_a and Y_b into a shared part, Y_S and unshared parts for each, $Y_{\delta a}$ and $Y_{\delta b}$.

$$\begin{aligned}\text{Cov}[Y_S + Y_{\delta a}, Y_S + Y_{\delta b}|T] &= \text{Var}[Y_S|T] = \frac{\theta}{2}\tau_{a+b}m_2. \\ E[\frac{\theta}{2}\tau_{a+b}m_2] &= \frac{\theta}{2}m_2\text{E}[\tau_{a+b}]\end{aligned}$$

$$\text{Cov}[E[Y_a|T], E[Y_b|T]] = \text{Cov}[\frac{\theta}{2}Tm_1, \frac{\theta}{2}Tm_1] = \left(\frac{\theta}{2}m_1\right)^2 \text{Var}[T_{MRC A}]$$

Therefore we have

$$\text{Cov}[Y_a, Y_b] = \frac{\theta}{2}m_2\text{E}[\tau_{a+b}] + \left(\frac{\theta}{2}m_1\right)^2 \text{Var}[T_{MRC A}]. \quad (20)$$

1.3 Calculation of moments

1.3.1 Low mutation rate approximation

The main utility of moment generating functions is to calculate moments. We showed how the distribution of trait values converges to a normal distribution when there are a large number of sites with small effect per mutation relative to the rate of mutation. In order to see how much deviation there is from this normal model we can calculate moments and compare them to the moments of the multivariate normal distribution.

At first we will do this using the same low mutation rate approximation that was used for the normal limit. The moment generation function of the trait values can be approximated as

$$\varphi_{\mathbf{Y}}(\mathbf{k}) \approx \left[1 + \sum_{\omega \in \Omega} E[T_{\omega}] \frac{\theta}{2} \left(\psi \left(\sum_{a \in \omega} k_a \right) - 1 \right) + O(\theta^2) \right]^L. \quad (21)$$

This expression ignores any terms of the genealogy moment generation function shown in equation ?? above order one. These terms are not written because they contain products of mutation rates and coalescent times of order two and greater. Removing these terms means that equation 21 is free of terms giving the expectations of products such as $E[\frac{\theta}{2} t_{\omega_1} \frac{\theta}{2} t_{\omega_2}]$. We can use equation (21) to express moments of the trait distribution in terms of expected branch lengths calculated from coalescent models.

1.3.2 The second moment and variance

We can use the low mutation rate approximation to the moment generating function (equation 21) to calculate moments of the distribution of trait values. We'll start by calculating the first and second moments and compare these to the mean and variance of the normal distribution. We start, as we did in deriving the normal distribution, by substituting the Taylor series of the mutational mgf.

$$\varphi_{\mathbf{Y}}(\mathbf{k}) \approx \left[1 + \sum_{\omega \in \Omega} E[t_{\omega}] \frac{\theta}{2} \left(m_1 \sum_{a \in \omega} k_a + \frac{m_2}{2!} \left(\sum_{a \in \omega} k_a \right)^2 + \frac{m_3}{3!} \left(\sum_{a \in \omega} k_a \right)^3 + \frac{m_4}{4!} \left(\sum_{a \in \omega} k_a \right)^4 \dots \right) \right]^L \quad (22)$$

We can expand this out using multinomial coefficients to get

$$\begin{aligned} \varphi_{\mathbf{Y}}(\mathbf{k}) \approx & 1 + L \frac{\theta}{2} \sum_{\omega \in \Omega} E[t_{\omega}] \left(m_1 \sum_{a \in \omega} k_a + \frac{m_2}{2} \left(\sum_{a \in \omega} k_a \right)^2 + \dots \right) \\ & + \frac{L(L-1)}{2} \left(\frac{\theta}{2} \right)^2 \sum_{\omega \in \Omega} E[t_{\omega}]^2 \left(m_1 \sum_{a \in \omega} k_a + \frac{m_2}{2} \left(\sum_{a \in \omega} k_a \right)^2 + \dots \right)^2 \\ & + L(L-1) \left(\frac{\theta}{2} \right)^2 \sum_{\omega_1, \omega_2 \in \Omega} E[t_{\omega_1}] E[t_{\omega_2}] \left(m_1 \sum_{a \in \omega_1} k_a + \dots \right) \left(m_1 \sum_{a \in \omega_2} k_a + \dots \right) + \dots \end{aligned} \quad (23)$$

The first coefficient is $\binom{L}{L-1,1,\mathbf{0}}$, the second is $\binom{L}{L-2,2,\mathbf{0}}$, and the third is $\binom{L}{L-2,1,1,\mathbf{0}}$. This is probably not the best way to think about the combinatorics, but it will do for now. To calculate

the moments of this distribution one takes the partial derivatives of the mgf and sets the dummy variables to zero.

$$E[Y_1^{r_1} \dots Y_n^{r_n}] = \frac{\partial^{r_1+\dots+r_n}}{\partial k_1^{r_1} \dots \partial k_n^{r_n}} \varphi_{\mathbf{Y}}(\mathbf{k}) \Big|_{\mathbf{k}=0} \quad (24)$$

Using this to calculate the first moment of the trait distribution we get

$$E[Y_a] \approx L \frac{\theta}{2} m_1 \sum_{\omega: a \in \omega} E[t_\omega]. \quad (25)$$

The second moment is more complicated because there are k_a^2 terms in all three lines of equation 23.

$$\begin{aligned} E[Y_a^2] &\approx L \frac{\theta}{2} m_2 \sum_{\omega: a \in \omega} E[t_\omega] \\ &+ \frac{L(L-1)}{2} \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega: a \in \omega} 2E[t_\omega]^2 \\ &+ L(L-1) \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega_1, \omega_2: a \in \omega_1, \omega_2} 2E[t_{\omega_1}]E[t_{\omega_2}] \end{aligned} \quad (26)$$

Terms with $(\frac{\theta}{2})^2$ are kept because they also include a second order term of L in front of them. We can now calculate the variance using $Var[Y] = E[Y^2] - E[Y]^2$. The squared first moment can be written as

$$\begin{aligned} \left(L \frac{\theta}{2} m_1 \sum_{\omega: a \in \omega} E[t_\omega] \right)^2 &= L^2 \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega: a \in \omega} E[t_\omega]^2 \\ &+ L^2 \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega_1, \omega_2: a \in \omega_1, \omega_2} E[t_{\omega_1}]E[t_{\omega_2}]. \end{aligned} \quad (27)$$

Subtracting this from the second moment gives

$$\begin{aligned} Var[Y_a] &\approx L \frac{\theta}{2} m_2 \sum_{\omega: a \in \omega} E[t_\omega] \\ &- L \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega: a \in \omega} E[t_\omega]^2 \\ &- 2L \left(\frac{\theta}{2}\right)^2 m_1^2 \sum_{\omega_1, \omega_2: a \in \omega_1, \omega_2} E[t_{\omega_1}]E[t_{\omega_2}] \\ &= L \frac{\theta}{2} m_2 \sum_{\omega: a \in \omega} E[t_\omega] - L \left(\frac{\theta}{2} m_1 \sum_{\omega: a \in \omega} E[t_\omega]\right)^2 \\ &= L \frac{\theta}{2} m_2 E[t_{MRCa}] - L \left(\frac{\theta}{2} m_1 E[t_{MRCa}]\right)^2 \\ &\approx L \frac{\theta}{2} m_2 E[t_{MRCa}] \end{aligned} \quad (28)$$

The $(\frac{\theta}{2})^2$ terms are not only first order in L so they can be ignored. This is the same variance as in the normal distribution because the limit takes care of these terms automatically.

1.3.3 The fourth moment

Due to the large number of terms I only derive the fourth moment of the trait value distribution for the case when the mean mutational effect is zero. Having the mean equal to zero is also helpful when comparing against a normal distribution because higher order moments of the normal distribution are easy to calculate when the mean is zero. The terms of (22) that will appear in the fourth moment after we apply (24) are

$$L \left(\frac{\theta}{2} \right) \frac{m_4}{24} \sum_{\omega: a \in \omega} E[t_\omega] \left(\sum_{a \in \omega} k_a \right)^4$$

for the fourth moment along one branch,

$$\binom{L}{L-2, 2, \mathbf{0}} \left(\frac{\theta}{2} \right)^2 \left(\frac{m_2}{2} \right)^2 24 \sum_{\omega: a \in \omega} E[t_\omega]^2 \left(\sum_{a \in \omega} k_a \right)^4$$

for the second moment of the same branch chosen twice, and

$$\binom{L}{L-2, 1, 1, \mathbf{0}} \left(\frac{\theta}{2} \right)^2 \left(\frac{m_2}{2} \right)^2 24 \sum_{\omega_1, \omega_2: a \in \omega_1, \omega_2} E[t_{\omega_1}] E[t_{\omega_2}] \left(\sum_{a \in \omega_1} k_a \right)^2 \left(\sum_{a \in \omega_2} k_a \right)^2$$

for the second moments on two different branches. Taking the fourth derivatives of these in terms of the desired branch we get

$$\begin{aligned} E[Y_a^4] &= L \frac{\theta}{2} m_4 E[T_{MRC A}] \\ &+ \frac{L(L-1)}{2} \left(\frac{\theta}{2} \right)^2 \left(\frac{m_2}{2} \right)^2 24 \sum_{\omega: a \in \omega} E[t_\omega]^2 \\ &+ L(L-1) \left(\frac{\theta}{2} \right)^2 \left(\frac{m_2}{2} \right)^2 24 \sum_{\omega_1, \omega_2: a \in \omega_1, \omega_2} E[t_{\omega_1}] E[t_{\omega_2}] \\ &= L \frac{\theta}{2} m_4 E[T_{MRC A}] + 3L(L-1) \left(\frac{\theta}{2} m_2 \sum_{\omega: a \in \omega} E[t_\omega] \right)^2 x \\ &= L \frac{\theta}{2} m_4 E[T_{MRC A}] + 3L(L-1) \left(\frac{\theta}{2} m_2 E[T_{MRC A}] \right)^2 \\ &\approx L \frac{\theta}{2} m_4 E[T_{MRC A}] + 3 \left(L \frac{\theta}{2} m_2 E[T_{MRC A}] \right)^2 \end{aligned} \tag{29}$$

1.3.4 Kurtosis

Since we know that the variance in trait values is very close under the low mutation rate model as in the normal limit, we might next compare kurtosis which provides a measure of the tailedness of the distribution. The kurtosis is defined as

$$Kurt[X] = \frac{E[(X - E[X])^4]}{(E[(X - E[X])^2])^2}. \tag{30}$$

This is the fourth central moment divided by the variance. For ease of calculation, we'll examine this in the case where the mean mutation effect (and therefore trait value) is zero. If we plug (28) and (29) into the expression for the kurtosis we get

$$\begin{aligned}
Kurt[Y_a] &= \frac{L \frac{\theta}{2} m_4 E[T_{MRCa}]}{(L \frac{\theta}{2} m_2 E[T_{MRCa}])^2} + \frac{3L(L-1) \left(\frac{\theta}{2} m_2 E[T_{MRCa}] \right)^2}{(L \frac{\theta}{2} m_2 E[T_{MRCa}])^2} \\
&= \frac{m_4}{L \frac{\theta}{2} m_2^2 E[T_{MRCa}]} + \frac{3(L^2 - L)}{L^2} \\
&= \frac{Kurt[M]}{L \frac{\theta}{2} E[T_{MRCa}]} + 3 \left(1 - \frac{1}{L} \right). \tag{31}
\end{aligned}$$

What does this mean? The normal approximation will have a kurtosis of 3, so having a finite number of loci will act to decrease the kurtosis of the trait distribution. However, it seems this decrease will only be very slight. The term on the left shows that the kurtosis is increased by the ratio of the kurtosis of the mutational distribution relative to the expected number of segregating sites affecting the trait.

1.3.5 Cokurtosis

Similarly to kurtosis, we can calculate another fourth order moment called the cokurtosis which measures the propensity of extreme values to occur together in the joint distribution. The cokurtosis is defined as

$$Cokurt(X, X, Y, Y) = \frac{E[(X - E[X])^2(Y - E[Y])^2]}{\sigma_X^2 \sigma_Y^2}. \tag{32}$$

We will only look at the balanced version of this for now. As with covariance, the kurtosis of a random variable is its cokurtosis with itself. Courtesies may be of interest residual signal of individual genealogies may cause individual to share extreme trait values due to shared ancestry. We calculate the cokurtosis by first calculating $E[Y_a^2 Y_b^2]$ in the same way as was done for the kurtosis by considering different terms of (22). One is the fourth moment term for branches containing both a and b :

$$L \frac{\theta}{2} \frac{m_4}{4!} \sum_{\omega: a, b \in \omega} E[t_\omega] \left(\sum_{d \in \omega} k_d \right)^4.$$

Another is the second moment terms for pairs of branches both containing a and b :

$$\frac{L(L-1)}{2} \left(\sum_{\omega: a, b \in \omega} E[t_\omega] \frac{\theta}{2} \frac{m_2}{2} \left(\sum_{d \in \omega} k_d \right)^2 \right).$$

Let $\Omega_{a/b}$ be the set of branches containing a but not b and Ω_{a+b} be the set of branches containing both a and b . We also need to consider terms from $\Omega_{a/b} \times \Omega_b$ and $\Omega_{b/a} \times \Omega_a$. These are

$$\begin{aligned} & L(L-1) \left(\sum_{\omega_1 \in \Omega_{a/b}} E[t_{\omega_1}] \frac{\theta}{2} \frac{m_2}{2} \left(\sum_{d_1 \in \omega} k_{d_1} \right)^2 \right) \left(\sum_{\omega_2 \in \Omega_b} E[t_{\omega_2}] \frac{\theta}{2} \frac{m_2}{2} \left(\sum_{d_2 \in \omega} k_{d_2} \right)^2 \right) \\ & + L(L-1) \left(\sum_{\omega_1 \in \Omega_{b/a}} E[t_{\omega_1}] \frac{\theta}{2} \frac{m_2}{2} \left(\sum_{d_1 \in \omega} k_{d_1} \right)^2 \right) \left(\sum_{\omega_2 \in \Omega_a} E[t_{\omega_2}] \frac{\theta}{2} \frac{m_2}{2} \left(\sum_{d_2 \in \omega} k_{d_2} \right)^2 \right) \\ & - L(L-1) \left(\sum_{\omega_1 \in \Omega_{b/a}} E[t_{\omega_1}] \frac{\theta}{2} \frac{m_2}{2} \left(\sum_{d_1 \in \omega} k_{d_1} \right)^2 \right) \left(\sum_{\omega_2 \in \Omega_{a/b}} E[t_{\omega_2}] \frac{\theta}{2} \frac{m_2}{2} \left(\sum_{d_2 \in \omega} k_{d_2} \right)^2 \right) \end{aligned}$$

The last term is subtracted because $\Omega_{a/b} \times \Omega_{b/c}$ gets counted twice. Taking the appropriate derivatives we get

$$\begin{aligned} E[Y_a^2 Y_b^2] &= L \frac{\theta}{2} m_4 \sum_{\omega \in \Omega_{a+b}} E[t_\omega] \\ &+ 3L(L-1) \left(\sum_{\omega \in \Omega_{a+b}} \frac{\theta}{2} m_2 E[t_\omega] \right)^2 \\ &+ L(L-1) \left(\sum_{\omega_1 \in \Omega_{a/b}} E[t_{\omega_1}] \frac{\theta}{2} m_2 \right) \left(\sum_{\omega_2 \in \Omega_b} E[t_{\omega_2}] \frac{\theta}{2} m_2 \right) \\ &+ L(L-1) \left(\sum_{\omega_1 \in \Omega_{b/a}} E[t_{\omega_1}] \frac{\theta}{2} m_2 \right) \left(\sum_{\omega_2 \in \Omega_a} E[t_{\omega_2}] \frac{\theta}{2} m_2 \right) \\ &- L(L-1) \left(\sum_{\omega_1 \in \Omega_{b/a}} E[t_{\omega_1}] \frac{\theta}{2} m_2 \right) \left(\sum_{\omega_2 \in \Omega_{a/b}} E[t_{\omega_2}] \frac{\theta}{2} m_2 \right) \end{aligned}$$

One way to simplify this is to notice that the last three lines almost sum over $\Omega_a \times \Omega_b$ except that they are missing the product over terms where both branches contain both a and b . If we subtract one of the three terms from the second line and add it here we get

$$\begin{aligned} E[Y_a^2 Y_b^2] &= L \frac{\theta}{2} m_4 \sum_{\omega \in \Omega_{a+b}} E[t_\omega] \\ &+ 2L(L-1) \left(\sum_{\omega \in \Omega_{a+b}} \frac{\theta}{2} m_2 E[t_\omega] \right)^2 \\ &+ L(L-1) \left(\sum_{\omega_1 \in \Omega_a} E[t_{\omega_1}] \frac{\theta}{2} m_2 \right) \left(\sum_{\omega_2 \in \Omega_b} E[t_{\omega_2}] \frac{\theta}{2} m_2 \right), \end{aligned}$$

which can be rewritten as

$$E[Y_a^2 Y_b^2] = L \frac{\theta}{2} m_4 E[\tau_{a+b}] + 2L(L-1) \left(\frac{\theta}{2} m_2 E[\tau_{a+b}] \right)^2 + L(L-1) \left(\frac{\theta}{2} m_2 E[T_{MRCA}] \right)^2. \quad (33)$$

Since $E[\tau_a]$ is the expected length of branches in the entire genealogy containing a , this is also the same as $E[T_{MRCA}]$.

1.3.6 Additional moments

We also calculate here some additional moments that have less clear interpretations but are useful later on when calculating the expected population kurtosis. The first of these is $E[Y_a^3 Y_b]$. The terms of (22) that will appear in this are

$$L \left(\frac{\theta}{2} \right) \frac{m_4}{24} 4k_a^3 k_b \sum_{\omega: a, b \in \omega} E[t_\omega]$$

and

$$L(L-1) \left(\frac{\theta}{2} \frac{m_2}{2} \right)^2 k_a^2 \times 2k_a k_b \left(\sum_{\omega: a, b \in \omega} E[t_\omega] \right) \left(\sum_{\omega: a \in \omega} E[t_\omega] \right).$$

This ultimately gives

$$E[Y_a^3 Y_b] = L \frac{\theta}{2} m_4 E[\tau_{a+b}] + 3L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[T_{MRCA}] E[\tau_{a+b}]. \quad (34)$$

The next fourth moment of interest is $E[Y_a^2 Y_b Y_c]$. The terms of (22) are

$$L \frac{\theta}{2} \frac{m_4}{24} 12k_a^2 k_b k_c \sum_{\omega: a, b, c \in \omega} E[t_\omega],$$

$$L(L-1) \left(\frac{\theta}{2} \frac{m_2}{2} \right)^2 k_a^2 \times 2k_b k_c \left(\sum_{\omega: a \in \omega} E[t_\omega] \right) \left(\sum_{\omega: b, c \in \omega} E[t_\omega] \right),$$

and

$$L(L-1) \left(\frac{\theta}{2} \frac{m_2}{2} \right)^2 2k_a k_b \times 2k_a k_c \left(\sum_{\omega: a, b \in \omega} E[t_\omega] \right) \left(\sum_{\omega: a, c \in \omega} E[t_\omega] \right).$$

Taking the appropriate derivatives of these gives

$$L \frac{\theta}{2} m_4 E[\tau_{a+b+c}] + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[T_{MRCA}] E[\tau_{b+c}] + 2L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}] E[\tau_{a+c}]. \quad (35)$$

Individuals in the population are exchangeable as long as it is not structured. The pairwise expected shared branch lengths are in that case all equal and we can write (35) as

$$L \frac{\theta}{2} m_4 E[\tau_{a+b+c}] + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[T_{MRCA}] E[\tau_{a+b}] + 2L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}]^2. \quad (36)$$

Where τ_{a+b} just refers to the expected shared branch length of any two individuals in the population. The final moment we'll look at is $E[Y_a Y_b Y_c Y_d]$ which has relevant terms

$$L \frac{\theta}{2} \frac{m_4}{24} 24k_a k_b k_c k_d \sum_{\omega: a, b, c, d \in \omega} E[t_\omega],$$

$$L(L-1) \left(\frac{\theta}{2} \frac{m_2}{2} \right)^2 2k_a k_b \times 2k_c k_d \left(\sum_{\omega: a, b \in \omega} E[t_\omega] \right) \left(\sum_{\omega: c, d \in \omega} E[t_\omega] \right),$$

$$L(L-1) \left(\frac{\theta}{2} \frac{m_2}{2} \right)^2 2k_a k_c \times 2k_b k_d \left(\sum_{\omega: a, c \in \omega} E[t_\omega] \right) \left(\sum_{\omega: b, d \in \omega} E[t_\omega] \right),$$

and

$$L(L-1) \left(\frac{\theta}{2} \frac{m_2}{2} \right)^2 2k_a k_d \times 2k_b k_c \left(\sum_{\omega: a, d \in \omega} E[t_\omega] \right) \left(\sum_{\omega: b, c \in \omega} E[t_\omega] \right).$$

When the appropriate fourth order partial derivatives are taken of this we get

$$\begin{aligned} & L \frac{\theta}{2} m_4 E[\tau_{a+b+c+d}] \\ & + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}] E[\tau_{c+d}] \\ & + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+c}] E[\tau_{b+d}] \\ & + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+d}] E[\tau_{b+c}]. \end{aligned}$$

We can again simplify this expression for populations with exchangeable individuals. This gives

$$L \frac{\theta}{2} m_4 E[\tau_{a+b+c+d}] + 3L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}]^2. \quad (37)$$

1.3.7 Comparison to normal distribution

The Isserlis theorem states that the moments of a multivariate normal random vector with mean zero is given by

$$E[X_1 X_2 \dots X_n] = \sum \prod E[X_i X_j]. \quad (38)$$

The sum is over all possible ways of dividing the n variables up into pairs, and the product is over all pairwise expectations in that partition.

1.3.8 General moments

Unfortunately, it does not seem possible to write down a simple solution to (24) that would provide an expression for an arbitrary moment of the distribution of trait values. For an order n term one would need to sum over all orders of expected branch length terms that would appear in the expansion of (22). For each order of branch lengths one would need to sum over all possible sets of branch lengths and moments for those branch lengths such that the desired product of dummy variables raised to the correct powers could be produced. Furthermore, one would need to calculate the coefficient for this product of dummy variables.

The one positive aspect about calculating general moments under the low mutation rate approximation is that they only depend on the expected branch lengths.

1.4 Population moments

1.4.1 Population variance

The expected variance of breeding values in the population, and what in this model would be considered the genetic variance is

$$E[V_g] = E \left[\frac{1}{N} \sum \left(Y_i - \frac{\sum Y_j}{N} \right)^2 \right]. \quad (39)$$

By assuming that all the individuals in the sample are exchangeable, we can expand the inside to get

$$\begin{aligned} E \left[\left(Y_i - \frac{\sum Y_j}{N} \right)^2 \right] &= E \left[Y_i^2 - 2Y_i \frac{\sum Y_j}{N} + \left(\frac{\sum Y_j}{N} \right)^2 \right] \\ &= E[Y_i^2] - \frac{2}{N} E[Y_i^2] - \frac{2(N-1)}{N} E[Y_i Y_j] + \frac{1}{N} E[Y_i^2] + \frac{N-1}{N} E[Y_i Y_j] \\ &= \frac{N-1}{N} (E[Y_i^2] - E[Y_i Y_j]). \end{aligned}$$

Thus when N is even moderately large

$$E[V_g] = E[Y_i^2] - E[Y_i Y_j]. \quad (40)$$

When $m_1 = 0$, $E[Y_i^2] \approx L \frac{\theta}{2} m_2 E[T_{MRC A}]$ and $E[Y_i Y_j] \approx L \frac{\theta}{2} m_2 (E[T_{MRC A}] - E[\tau_{i,j}])$. The expected genetic variance in the population is then proportional to the expected pairwise coalescent time as we would expect.

1.4.2 Variance of population variance

The variance of the population variance over evolutionary realizations is another useful quantity. In the infinitesimal limit for a sufficiently large population, the variance of the variance will be negligible relative to the variance itself. That is, the coefficient of variation of the variance will approach zero. Here I will derive the variance of the variance, from which one could compute the coefficient of variation.

$$\begin{aligned} \text{Var}[V_g] &= \text{Var} \left[\frac{1}{N} \sum \left(Y_i - \frac{\sum Y_j}{N} \right)^2 \right] \\ &= \frac{1}{N^2} \left(\sum \text{Var}[(Y_i - \frac{\sum Y_j}{N})^2] + 2 \sum_{i \neq j} \text{Cov}[(Y_i - \frac{\sum Y_k}{N})^2, (Y_j - \frac{\sum Y_k}{N})^2] \right). \end{aligned} \quad (41)$$

The first term can be ignored if the population size is large. The second term is determined by the covariance in the squared deviations from the mean. The covariance can be calculated by considering its individual parts.

$$\begin{aligned}
\text{Cov}[Y_i^2, Y_j^2] &\approx -2\text{Cov}[Y_i^2, Y_j^2] \\
\text{Cov}[Y_i^2, -2Y_j \frac{\sum Y_k}{N}] &\approx -2\text{Cov}[Y_i^2, Y_j Y_k] \\
\text{Cov}[Y_i^2, \left(\frac{\sum Y_k}{N}\right)^2] &\approx \text{Cov}[Y_i^2, Y_j Y_k] \\
\text{Cov}[-2Y_i \left(\frac{\sum Y_k}{N}\right), Y_j^2] &\approx -2\text{Cov}[Y_i^2, Y_j Y_k] \\
\text{Cov}[-2Y_i \left(\frac{\sum Y_k}{N}\right), -2Y_j \left(\frac{\sum Y_k}{N}\right)] &\approx 4\text{Cov}[Y_i Y_j, Y_k Y_l] \\
\text{Cov}[-2Y_i \left(\frac{\sum Y_k}{N}\right), \left(\frac{\sum Y_k}{N}\right)] &\approx -2\text{Cov}[Y_i Y_j, Y_k Y_l] \\
\text{Cov}[\left(\frac{\sum Y_k}{N}\right)^2, Y_j^2] &\approx \text{Cov}[Y_j^2, Y_i Y_k] \\
\text{Cov}[\left(\frac{\sum Y_k}{N}\right)^2, -2Y_j \left(\frac{\sum Y_k}{N}\right)] &\approx -2\text{Cov}[Y_i Y_j, Y_k Y_l] \\
\text{Cov}[\left(\frac{\sum Y_k}{N}\right)^2, \left(\frac{\sum Y_k}{N}\right)^2] &\approx \text{Cov}[Y_i Y_j, Y_k Y_l]
\end{aligned}$$

Combining all these gives

$$\text{Cov}[(Y_i - \frac{\sum Y_k}{N})^2, (Y_j - \frac{\sum Y_k}{N})^2] = \text{E}[Y_i^2 Y_j^2] - \text{E}[Y_i^2]^2 - 2\text{E}[Y_i^2 Y_j Y_k] + 2\text{E}[Y_i^2] \text{E}[Y_j Y_k] + \text{E}[Y_i Y_j Y_k Y_l] - \text{E}[Y_i Y_j]^2. \quad (42)$$

After calculating the various moments included in this sum and using the low mutation rate approximation we get

$$L \frac{\theta}{2} m_4 \left(\frac{\text{E}[T_{2,4}]}{9} + \frac{\text{E}[T_{3,4}]}{6} \right). \quad (43)$$

With this, this coefficient of variation of the variance is

$$\sqrt{\frac{\kappa}{L \frac{\theta}{2} \text{E}[T_{2,2}]}} \sqrt{\frac{\frac{\text{E}[T_{2,4}]}{9} + \frac{\text{E}[T_{3,4}]}{6}}{\text{E}[T_{2,2}]}}. \quad (44)$$

The first term depends on the genetic architecture while the second depends on demography. From the first we can see that the coefficient of variation of variance decreases with the square root of the expected number of trait affecting separating two haplotypes.

1.4.3 Population fourth central moment

A good measure for the deviation from a normal distribution is the kurtosis, which is the ratio of the fourth central moment to the square of the variance. For the normal distribution this value is three. The kurtosis measures the propensity of a distribution to produce outliers. By calculating the expected kurtosis for a given genetic architecture and population history we could assess the propensity of those parameters to produce trait distributions (at the population level) with frequent outliers.

However, the expected kurtosis in the population is a quotient and therefore very annoying to calculate. We could approximate the kurtosis by taking the expected fourth moment divided by the expectation of the square of the variance. A better approximation could be obtained through a Taylor expansion, but that would involve calculation of trait moments of order eight and greater.

Instead we will calculate the expected fourth central moment relative to the expected fourth moment in a normal distribution.

$$E[M_{4,Y}] = E \left[\frac{1}{N} \sum \left(Y_i - \frac{\sum Y_j}{N} \right)^4 \right]. \quad (45)$$

Examining the sum inside the expectation we see that

$$\begin{aligned} E \left[\left(Y_i - \frac{\sum Y_j}{N} \right)^4 \right] &= E[Y_i^4] - 4E \left[Y_i^3 \frac{\sum Y_j}{N} \right] + 6E \left[Y_i^2 \left(\frac{\sum Y_j}{N} \right)^2 \right] - 4E \left[Y_i \left(\frac{\sum Y_j}{N} \right)^3 \right] + \left(\frac{\sum Y_j}{N} \right)^4 \\ &= E[Y_i^4] - \frac{4}{N} \sum_j E[Y_i^3 Y_j] + \frac{6}{N^2} \sum_{j,k} E[Y_i^2 Y_j Y_k] \\ &\quad - \frac{4}{N^3} \sum_{j,k,l} E[Y_i Y_j Y_k Y_l] + \frac{1}{n^4} \sum_{j,k,l,d} E[Y_j Y_k Y_l Y_d]. \end{aligned} \quad (46)$$

In calculating these expectations we have to remember that the value depends only on the number of times each variable appears in the expectation. That is, $E[Y_1^2 Y_2 Y_3]$ is equivalent to $E[Y_1 Y_2^3 Y_3]$ as long as all individuals in the population are exchangeable. The resulting expansion of (46) is therefore quite ugly. It can be simplified by only considering terms of order one. Other terms can be ignored since we are assuming there are a fair number of individuals in the population. This yields

$$\begin{aligned} E \left[\left(Y_i - \frac{\sum Y_j}{N} \right)^4 \right] &= E[Y_i^4] - \frac{4(n-1)}{n} E[Y_i^3 Y_j] + \frac{6(n-1)(n-2)}{n^2} E[Y_i^2 Y_j Y_k] \\ &\quad - \frac{4(n-1)(n-2)(n-3)}{n^3} E[Y_i Y_j Y_k Y_l] \\ &\quad + \frac{(n-1)(n-2)(n-3)(n-4)}{n^4} E[Y_j Y_k Y_l Y_d] + O(n^{-1}) \\ &\approx E[Y_i^4] - 4E[Y_i^3 Y_j] + 6E[Y_i^2 Y_j Y_k] - 3E[Y_i Y_j Y_k Y_l]. \end{aligned} \quad (47)$$

The first term, $E[Y_i^4]$ was derived in equation (29) as

$$L \frac{\theta}{2} m_4 E[T_{MRC A}] + 3L(L-1) \left(\frac{\theta}{2} m_2 E[T_{MRC A}] \right)^2.$$

The second term, $E[Y_i^3 Y_j]$ was derived in equation (34) as

$$L \frac{\theta}{2} m_4 E[\tau_{a+b}] + 3L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[T_{MRC A}] E[\tau_{a+b}].$$

The third term, $E[Y_i^2 Y_j Y_k]$ was derived in equation (35) as

$$L \frac{\theta}{2} m_4 E[\tau_{a+b+c}] + L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[T_{MRC A}] E[\tau_{a+b}] + 2L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}]^2.$$

The fourth term, $E[Y_i Y_j Y_k Y_l]$ was derived in equation (37) as

$$L \frac{\theta}{2} m_4 E[\tau_{a+b+c+d}] + 3L(L-1) \left(\frac{\theta}{2} m_2 \right)^2 E[\tau_{a+b}]^2.$$

Plugging these into (47) we get

$$\begin{aligned} E[M_{4,Y}] &\approx L \frac{\theta}{2} m_4 (E[T_{MRC A}] - 4E[\tau_{a+b}] + 6E[\tau_{a+b+c}] - 3E[\tau_{a+b+c+d}]) \\ &\quad + 3 \left(L \frac{\theta}{2} m_2 \right)^2 (E[T_{MRC A}] - E[\tau_{a+b}])^2. \end{aligned} \quad (48)$$

These expressions approximate $L(L-1)$ as L^2 as part of the low mutation rate approximation. The fourth central moment of a normal distribution is $3\sigma^4$. Using the expected population variance we get $3 \left(L \frac{\theta}{2} m_2 E[T_{MRC A}] - E[\tau_{a+b}] \right)^2$. It is clear from (48) that as $L \frac{\theta}{2}$ gets large the second term will dominate and the population will have the same expected fourth moment as a normal distribution. The ratio of the expected fourth moment to that under normality is

$$\frac{E[M_4]}{3 \left(L \frac{\theta}{2} m_2 E[T_{MRC A}] - E[\tau_{a+b}] \right)} \approx 3 + \frac{\text{Kurt}[M] (E[T_{MRC A}] - 4E[\tau_{a+b}] + 6E[\tau_{a+b+c}] - 3E[\tau_{a+b+c+d}])}{L \frac{\theta}{2} E[T_{MRC A}] - E[\tau_{a+b}]}.$$

This expression can be written in an easier to interpret form by noting that $E[\tau_{a+b}] = E[T_{MRC A}] - E[T_2]$, $E[\tau_{a+b+c}] = E[T_{MRC A}] - E[T_3]$, $E[\tau_{a+b+c+d}] = E[T_{MRC A}] - E[T_4]$, where T_i is the expected time it takes for i lineages to coalesce.

$$3 + \frac{\text{Kurt}[M] (4T_2 - 6T_3 + 3T_4)}{L \frac{\theta}{2} T_2^2}. \quad (49)$$

In a constant size population this would be $3 + \frac{\text{Kurt}[M]}{2L \frac{\theta}{2}}$

1.5 First order approximation to the population kurtosis

The expected kurtosis at the population level is given by

$$E[\text{Kurt}] = E \left[\frac{\frac{1}{N} \sum_i (Y_i - \bar{Y})^4}{\left(\frac{1}{N} \sum_i (Y_i - \bar{Y})^2 \right)^2} \right]. \quad (50)$$

A first order approximation to this is

$$\frac{E[\frac{1}{N} \sum_i (Y_i - \bar{Y})^4]}{E \left[\left(\frac{1}{N} \sum_i (Y_i - \bar{Y})^2 \right)^2 \right]}. \quad (51)$$

The numerator is just the expected fourth moment at the population level. The denominator is $E[V_p]^2 + \text{Var}[V_p]$ if V_p is the variance at the population level. Using the results in equations 43

and 48 the whole expression can be written as

$$\begin{aligned}
\frac{E\left[\frac{1}{N}\sum_i(Y_i - \bar{Y})^4\right]}{E\left[\left(\frac{1}{N}\sum_i(Y_i - \bar{Y})^2\right)^2\right]} &= \frac{3\left(L\frac{\theta}{2}E[T_{2,2}]\right)^2 + L\frac{\theta}{2}m_4\left(E[T_{4,4}] + \frac{1}{3}E[T_{3,4}] + \frac{2}{9}E[T_{2,4}]\right)}{\left(L\frac{\theta}{2}m_2E[T_{2,2}]\right)^2 + L\frac{\theta}{2}m_4\left(\frac{1}{9}E[T_{2,4}] + \frac{1}{6}E[T_{3,4}]\right)} \\
&= 3 + \frac{L\frac{\theta}{2}m_4\left(E[T_{4,4}] + \frac{1}{3}E[T_{3,4}] + \frac{2}{9}E[T_{2,4}] - 3\left(\frac{1}{9}E[T_{2,4}] + \frac{1}{6}E[T_{3,4}]\right)\right)}{\left(L\frac{\theta}{2}m_2E[T_{2,2}]\right)^2 + L\frac{\theta}{2}m_4\left(\frac{1}{9}E[T_{2,4}] + \frac{1}{6}E[T_{3,4}]\right)} \\
&= 3 + \frac{\kappa\left(E[T_{4,4}] - \frac{1}{6}E[T_{3,4}] - \frac{1}{9}E[T_{2,4}]\right)}{L\frac{\theta}{2}E[T_{2,2}]^2 + \kappa\left(\frac{1}{9}E[T_{2,4}] + \frac{1}{6}E[T_{3,4}]\right)} \\
&= 3 + \frac{\kappa\frac{E[T_{4,4}] - \frac{1}{6}E[T_{3,4}] - \frac{1}{9}E[T_{2,4}]}{E[T_{2,2}]}}{L\frac{\theta}{2}E[T_{2,2}] + \kappa\frac{\frac{1}{9}E[T_{2,4}] + \frac{1}{6}E[T_{3,4}]}{E[T_{2,2}]}}.
\end{aligned}$$

We can see from this that when $E[T_{4,4}] - \frac{1}{6}E[T_{3,4}] - \frac{1}{9}E[T_{2,4}] < 0$ the expected kurtosis might be less than three, and that when $E[T_{4,4}] - \frac{1}{6}E[T_{3,4}] - \frac{1}{9}E[T_{2,4}] > 0$ it would be greater than three. For the standard coalescent model $E[T_{4,4}] - \frac{1}{6}E[T_{3,4}] - \frac{1}{9}E[T_{2,4}] = 0$ so the expected kurtosis is about three.

1.5.1 Fourth population moment with a step change

To get a sense for the effect that demography can have on the expected fourth moment in the population we have to consider a particular population size history. In general, finding expressions for expected coalescent times beyond a sample size of two are quite difficult under nonequilibrium population histories. A relatively simple case is that of a single step change in population size. Let N_0 be the current effective population size, z be the number of generations in the past when the population size changes, and N_1 be the size it changes to. A somewhat nicer parameterization uses $b = \frac{z}{N_0}$ and $c = \frac{N_1}{N_0}$. The expected times to the most recent common ancestor for samples of size two, three, and four are

$$\begin{aligned}
E[T_2] &= N_0(1 - e^{-b} + ce^{-b}) \\
E[T_3] &= N_0\left(\frac{1}{6}e^{-3b}(1 - c) - \frac{3}{2}e^{-b}(1 - c) + \frac{4}{3}\right) \\
E[T_4] &= N_0\left(\frac{-1}{30}e^{-6b} + \frac{1}{3}e^{-3b} - \frac{9}{5}e^{-b} + \frac{3}{2} + c\left(\frac{1}{30}e^{-6b} - \frac{1}{3}e^{-3b} + \frac{9}{5}e^{-b}\right)\right).
\end{aligned} \tag{52}$$

Using these we can see that the scaling factor for the fourth moment due to demography is

$$Q = \frac{4E[T_2] - 6E[T_3] + 3E[T_4]}{E[T_2]} = \frac{(e^{2b} - e^{-2b})(1 - c)}{2(e^b - 1 + c)}. \tag{53}$$

Figure 1.5.1 shows how the expected excess in the fourth population moment depends on the parameters b and c of the population size change.

1.6 Differences in trait values

1.6.1 Kurtosis

The distribution of a single trait value Y is not very interesting because this quantity cannot be observed. Y gives the change in the trait value since the most recent common ancestor of the

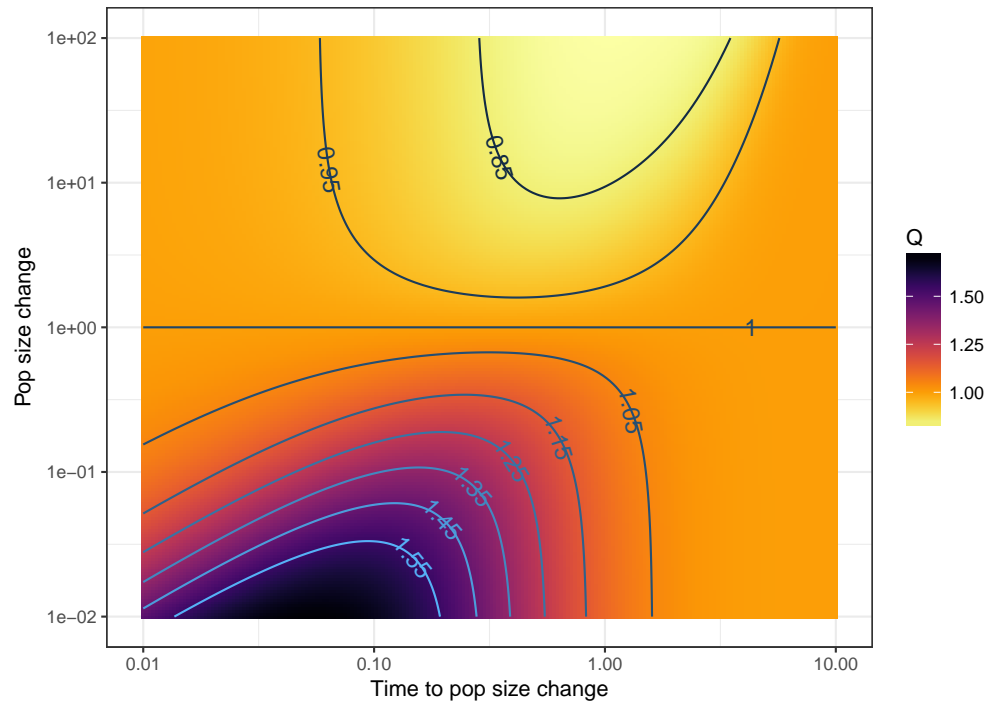


Figure 1: The scaling factor under different step population size changes.

sample, but since we don't know what that value is we can't measure Y in a sampled individual. What we can observe is the differences in trait values between individuals, $Y_a - Y_b$. The results for these differences are what one would expect given the distribution of the trait values themselves. For instance, the second and fourth moments look exactly like (28) and (29) if one substitutes twice the expected coalescent time for the time to the most recent common ancestor of the sample. One way to see this is by deriving the moment generating function for the difference in trait values between two individuals. Using the same arguments that were used to derive (10) we get

$$\begin{aligned} \varphi_{Y_a - Y_b}(k) &= \int \int \prod_{\omega_1 \in \Omega_{a/b}} e^{kY_{a,\omega_1}} P(Y_{a,\omega_1} = y_{a,\omega_1} | \mathbf{T} = \mathbf{t}) dy_{a,\omega_1} \\ &\times \int \prod_{\omega_2 \in \Omega_{b/a}} e^{-kY_{b,\omega_2}} P(Y_{b,\omega_2} = y_{b,\omega_2} | \mathbf{T} = \mathbf{t}) dy_{b,\omega_2} P(\mathbf{T} = \mathbf{t}) d\mathbf{t}. \end{aligned} \quad (54)$$

From this we have removed all branches where $Y_a - Y_b$ is necessarily zero. This includes branches that subtend both individuals and branches that subtend neither individual. From this it is clear, as expected, that the distribution of the trait difference between two individuals depends only on the distribution of coalescence times for those individuals. Using the same result for compound Poisson processes as before we get

$$\begin{aligned} \varphi_{Y_a - Y_b}(k) &= \int \prod_{\omega_1 \in \Omega_{a/b}} \exp\left(\frac{\theta}{2} t_{\omega_1} (\Phi(k) - 1)\right) \\ &\times \prod_{\omega_2 \in \Omega_{b/a}} \exp\left(\frac{\theta}{2} t_{\omega_2} (\Phi(-k) - 1)\right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t}. \end{aligned} \quad (55)$$

It is simple to use the low mutation rate approximation as in (22) to derive the second and fourth moments of the trait difference. These are:

$$E[(Y_a - Y_b)^2] = 2L \frac{\theta}{2} m_2 E[\tau_{a/b}]. \quad (56)$$

and

$$E[(Y_a - Y_b)^4] = 2L \frac{\theta}{2} m_4 E[\tau_{a/b}] + 12L(L-1) \left(\frac{\theta}{2} m_2 E[\tau_{a/b}]\right)^2. \quad (57)$$

The kurtosis is therefore

$$Kurt[Y_a - Y_b] = \frac{Kurt[M]}{2L \frac{\theta}{2} E[\tau_{a/b}]} + 3 \left(1 - \frac{1}{L}\right). \quad (58)$$

Again, the kurtosis depends on the kurtosis of the mutational distribution and the expected number of mutations affecting the trait. The kurtosis of the difference between individuals in different populations will be smaller than for individuals sampled within populations because the coalescence times will be greater for individuals samples from different populations.

1.6.2 Cokurtosis

Since the mean of a sample is uninformative, we have $n - 1$ meaningful observations in a sample of size n . Let Y_0 be the reference trait value and $Y_1 - Y_0, Y_2 - Y_0, \dots, Y_{n-1} - Y_0$ are the observed

trait differences. Let X be the random vector of these differences. We are now interested in the distribution of these differences. The moment generating function is

$$\varphi_{\mathbf{X}}(\mathbf{k}) = \int e^{\mathbf{k} \cdot \mathbf{x}} \int P(\mathbf{X} = \mathbf{x} | \mathbf{T} = \mathbf{t}) P(\mathbf{T} = \mathbf{t}) d\mathbf{t} d\mathbf{x}. \quad (59)$$

For the present we're going to look at $Cokurt[X_1, X_1, X_2, X_2]$, the cokurtosis between two trait differences in a sample. This should hopefully tell us something about the propensity of extreme trait values to be shared because of shared ancestry of the individuals. The mgf for this is

$$\int \int e^{k_1(y_a - y_0) + k_2(y_b - y_0)} P(Y_a - Y_0 = y_a - y_0, Y_b - Y_0 = y_b - y_0 | \mathbf{T} = \mathbf{t}) P(\mathbf{T} = \mathbf{t}) d\mathbf{y} d\mathbf{t}. \quad (60)$$

As before we can break this into different sections of the genealogy because the changes in trait values along these branches are independent conditional on T . Ignoring sets of branches where $Y_a - Y_0 = Y_b - Y_0 = 0$, the relevant branch sets are $(\Omega_{a/(b,0)}, \Omega_{(0+b)/a}, \Omega_{b/(0,a)}, \Omega_{(0+a)/b}, \Omega_{0/(a,b)}, \Omega_{(a+b)/0})$. Only considering these gives the following mgf

$$\begin{aligned} \varphi_{Y_a - Y_0, Y_b - Y_0}(k_1, k_2) &= \int \prod_{\omega \in \Omega_{a/(b,0)}} \exp\left(\frac{\theta}{2} t_\omega (\phi(k_1) - 1)\right) \\ &\times \prod_{\omega \in \Omega_{(0+b)/a}} \exp\left(\frac{\theta}{2} t_\omega (\phi(-k_1) - 1)\right) \\ &\times \prod_{\omega \in \Omega_{b/(0,a)}} \exp\left(\frac{\theta}{2} t_\omega (\phi(k_2) - 1)\right) \\ &\times \prod_{\omega \in \Omega_{(0+a)/b}} \exp\left(\frac{\theta}{2} t_\omega (\phi(-k_2) - 1)\right) \\ &\times \prod_{\omega \in \Omega_{0/(a,b)}} \exp\left(\frac{\theta}{2} t_\omega (\phi(-k_1 - k_2) - 1)\right) \\ &\times \prod_{\omega \in \Omega_{(a+b)/0}} \exp\left(\frac{\theta}{2} t_\omega (\phi(k_1 + k_2) - 1)\right) P(\mathbf{T} = \mathbf{t}) d\mathbf{t} \end{aligned}$$

Application of the low mutation rate approximation and assuming that m_1 and m_3 are zero gives

$$\begin{aligned}
& (1 + \sum_{\omega \in \Omega_{a/(0,b)}} E[t_\omega] \frac{\theta}{2} \left[\frac{m_2}{2} k_1^2 + \frac{m_4}{24} k_1^4 \right] \\
& + \sum_{\omega \in \Omega_{(0+b)/a}} E[t_\omega] \frac{\theta}{2} \left[\frac{m_2}{2} k_1^2 + \frac{m_4}{24} k_1^4 \right] \\
& + \sum_{\omega \in \Omega_{b/(0,a)}} E[t_\omega] \frac{\theta}{2} \left[\frac{m_2}{2} k_2^2 + \frac{m_4}{24} k_2^4 \right] \\
& + \sum_{\omega \in \Omega_{(0+a)/b}} E[t_\omega] \frac{\theta}{2} \left[\frac{m_2}{2} k_2^2 + \frac{m_4}{24} k_2^4 \right] \\
& + \sum_{\omega \in \Omega_{0/(a,b)}} E[t_\omega] \frac{\theta}{2} \left[\frac{m_2}{2} (k_1 + k_2)^2 + \frac{m_4}{24} (k_1 + k_2)^4 \right] \\
& + \sum_{\omega \in \Omega_{(a+b)/0}} E[t_\omega] \frac{\theta}{2} \left[\frac{m_2}{2} (k_1 + k_2)^2 + \frac{m_4}{24} (k_1 + k_2)^4 \right])^L.
\end{aligned}$$

If we ignore terms that won't contribute to the cokurtosis and do some grouping we get

$$\begin{aligned}
& (1 + \sum_{\omega \in \Omega_{0/a}} E[t_\omega] \frac{\theta}{2} \frac{m_2}{2} k_1^2 \\
& + \sum_{\omega \in \Omega_{a/0}} E[t_\omega] \frac{\theta}{2} \frac{m_2}{2} k_1^2 \\
& + \sum_{\omega \in \Omega_{0/b}} E[t_\omega] \frac{\theta}{2} \frac{m_2}{2} k_2^2 \\
& + \sum_{\omega \in \Omega_{b/0}} E[t_\omega] \frac{\theta}{2} \frac{m_2}{2} k_2^2 \\
& + \sum_{\omega \in \Omega_{0/(a,b)}} E[t_\omega] \frac{\theta}{2} \left[\frac{m_2}{2} 2k_1 k_2 + \frac{m_4}{24} k_1^2 k_2^2 \right] \\
& + \sum_{\omega \in \Omega_{(a,b)/0}} E[t_\omega] \frac{\theta}{2} \left[\frac{m_2}{2} 2k_1 k_2 + \frac{m_4}{24} k_1^2 k_2^2 \right])^L.
\end{aligned}$$

These sums over internal branches can be interpreted in terms of coalescence times.

$$\begin{aligned}
& (1 + 2E[\tau_{a/0}] \frac{\theta}{2} \frac{m_2}{2} k_1^2 \\
& + 2E[\tau_{b/0}] \frac{\theta}{2} \frac{m_2}{2} k_2^2 \\
& + E[\tau_{0/(a,b)}] \frac{\theta}{2} \left[\frac{m_2}{2} 2k_1 k_2 + \frac{m_4}{24} 6k_1^2 k_2^2 \right] \\
& + E[\tau_{(a,b)/0}] \frac{\theta}{2} \left[\frac{m_2}{2} 2k_1 k_2 + \frac{m_4}{24} 6k_1^2 k_2^2 \right])^L.
\end{aligned}$$

y As before $E[\tau_{a/0}]$ is the expected coalescence time between individuals a and 0. $E[\tau_{0/(a,b)}]$ is the expected total branch length subtending individual 0 before this lineage coalesces with

either a or b . $E[\tau_{(a,b)/0}]$ is the expected total branch length subtending both a and b before either coalesces with 0. This is a much less common a genealogical quantity than the expected coalescent time, but it seems to make sense in this context. Taking the appropriate derivatives we get

$$\begin{aligned} E[Y_a - Y_0, Y_b - Y_0] &= 4L(L-1)E[\tau_{a/0}]E[\tau_{b/0}] \left(\frac{\theta}{2}\right)^2 m_2^2 \\ &\quad + 2L(L-1)(E[\tau_{0/(a,b)}] + E[\tau_{(a,b)/0}])^2 \left(\frac{\theta}{2}\right)^2 m_2^2 \\ &\quad + LE[\tau_{0/(a,b)}] \frac{\theta}{2} m_4 + LE[\tau_{(a,b)/0}] \frac{\theta}{2} m_4. \end{aligned} \quad (61)$$

The cokurtosis is then

$$\left(1 - \frac{1}{L}\right) \left(1 + \frac{(E[\tau_{0/(a,b)}] + E[\tau_{(a,b)/0}])^2}{2E[\tau_{a/0}]E[\tau_{b/0}]}\right) + \frac{Kurt[M] (E[\tau_{0/(a,b)}] + E[\tau_{(a,b)/0}])}{4L \frac{\theta}{2} E[\tau_{a/0}]E[\tau_{b/0}]}. \quad (62)$$

The cokurtosis of a multivariate normal distribution is $1 + 2\rho$ where ρ is the correlation between the two variables. With this in mind we can rewrite (61) as

$$\left(1 - \frac{1}{L}\right) \left(1 + 2 \frac{(E[\tau_{0/(a,b)}] + E[\tau_{(a,b)/0}])^2}{4E[\tau_{a/0}]E[\tau_{b/0}]}\right) + \frac{Kurt[M] (E[\tau_{0/(a,b)}] + E[\tau_{(a,b)/0}])}{4L \frac{\theta}{2} E[\tau_{a/0}]E[\tau_{b/0}]}. \quad (63)$$

Here, $\frac{(E[\tau_{0/(a,b)}] + E[\tau_{(a,b)/0}])^2}{4E[\tau_{a/0}]E[\tau_{b/0}]}$ makes sense as the correlation between $Y_a - Y_0$ and $Y_b - Y_0$. (61) therefore has the same form as the other kurtosis formulas where the excess kurtosis above the normal expectation is proportional to the mutational kurtosis.

1.7 Genealogy moment generating functions for structured populations

LOHSE *et al.* (2011) found general forms for the generating functions of genealogies. In a coalescent model where events i occur at rate λ_i , the moment generating function for the genealogy can be written as

$$\varphi_T[\mathbf{T}] = \frac{\sum_i \lambda_i \varphi_i[\mathbf{s}]}{\sum_i \lambda_i + \sum_{|\omega|=1} s_\omega}. \quad (64)$$

This is a convolution of the time to the first event with a mixture of times after the first event, where the mixture is over all possible next events. We can apply this equation to a generalized structured population model where, M is the number of demes, $m_{i,j}$ is the migration rate from deme i to deme j , η_i is the coalescent rate within deme i , and $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_M\}$ is the collection of lineages currently active in each deme. In this model, events occur at rate

$$\sum_{i=1}^M \binom{|\Omega_i|}{2} \eta_i + \sum_{(i,j:i \neq j)} m_{i,j} |\Omega_i|, \quad (65)$$

where the first term is due to coalescence events and the second term is due to migration events. We first define operations on Ω due to these events. A event of lineage ω migrating from deme i to deme j would be represented as

$$\Omega(i : -\omega, j : +\omega) := \{\Omega \setminus \{\Omega_i, \Omega_j\}\} \cup \{\Omega_j \setminus \omega, \Omega_i \cup \omega\}.$$

A coalescent event of lineages x and y in deme i would be written as

$$\Omega(i : a \cup b) := \{\Omega \setminus \Omega_i\} \cup \{\{\Omega_i \cup ab\} \setminus \{a, b\}\}$$

The overall generating function is then

$$\varphi_{\mathbf{T}}^{\Omega}(\mathbf{s}) = \left(\sum_{i=1}^M \binom{|\Omega_i|}{2} \eta_i + \sum_{(i,j):i \neq j} m_{i,j} |\Omega_i| - \sum_{i=1}^M \sum_{\omega \in \Omega_i} s_{\omega} \right)^{-1} \quad (66)$$

$$\times \left(\sum_{i=1}^M \eta_i \sum_{(a,b) \in \Omega_i: a \neq b} \varphi_{\mathbf{T}}^{\Omega(i:a \cup b)}(\mathbf{s}) + \sum_{(i,j):i \neq j} m_{i,j} \sum_{\omega \in \Omega_i} \varphi_{\mathbf{T}}^{\Omega(i:-\omega, j: +\omega)}(\mathbf{s}) \right). \quad (67)$$

b If there are N sample lineages, this gives a system of $M \binom{N}{2}^{-1}$ equations which is a fuckton.

1.8 Trait value moment generating functions for structured populations

To convert this into a moment generating function for branch lengths we would make the substitution specified by equation 10. This gives

$$\varphi_{\mathbf{Y}}^{\Omega}(\mathbf{k}) = \left(\sum_{i=1}^M \binom{|\Omega_i|}{2} \eta_i + \sum_{(i,j):i \neq j} m_{i,j} |\Omega_i| - \sum_{i=1}^M \sum_{\omega \in \Omega_i} \frac{\theta}{2} \left(\psi \left(\sum_{a \in \omega} k_a \right) - 1 \right) \right)^{-1} \quad (68)$$

$$\times \left(\sum_{i=1}^M \eta_i \sum_{(a,b) \in \Omega_i: a \neq b} \varphi_{\mathbf{Y}}^{\Omega(i:a \cup b)}(\mathbf{k}) + \sum_{(i,j):i \neq j} m_{i,j} \sum_{\omega \in \Omega_i} \varphi_{\mathbf{Y}}^{\Omega(i:-\omega, j: +\omega)}(\mathbf{k}) \right). \quad (69)$$

If we are only interested in calculating particular moments of this distribution we can simplify this expression substantially. For instance, if we want the second moment of k , then this depends only on the expected TMRCA. We will be taking the second derivative of φ with respect to an arbitrary k_i and also substituting zero for each k_i . For solving the recursion which we will then take the second derivative of and evaluate at zero, we define a function $g(\mathbf{n}, k)$. The recursion for this function is

$$g(\mathbf{n}, k) = \left(\sum_{i=1}^M \binom{n_i}{2} \eta(i) + \sum_{i=1}^M n_i \sum_{j=1}^M m(i, j) + \frac{\theta}{2} \frac{k^2}{2} m_2 \right)^{-1} \\ \times \left(\sum_{i=1}^M \binom{n_i}{2} \eta(i) g(c(i, \mathbf{n}), k) + \sum_{i=1}^M n_i \sum_{j=1}^M m(i, j) g(e(i, j, \mathbf{n}), k) \right). \quad (70)$$

In this, $c(i, \mathbf{n})$ is a coalescent operation that removes a lineage from deme i . and $e(i, j, \mathbf{n})$ is a migration operation that moves a lineage from deme i to deme j .

1.9 An example of the infinitesimal limit in panmictic populations

We look at an example from a panmictic population to see how the genealogy generates a correlation structure among the samples. In particular, we take limits corresponding to an infinitesimal model and show that this results in multivariate normal distribution.

Recall that the moment generating function of a mutational distribution $f(u)$ can be written as

$$\psi_U(k) = 1 + kE[U] + \frac{k^2}{2!}E[U^2] + \frac{k^3}{3!}E[U^3] + \dots$$

We are going to assume that the mean mutational effect is zero, and therefore $E[U^2] = \text{Var}[U] := \tau^2$. We then require that $\lim_{L \rightarrow \infty} \tau^2 L \rightarrow \sigma^2$ and $\lim_{L \rightarrow \infty} E[U^k]L \rightarrow 0$ for $k > 2$. This roughly corresponds to a situation where there are very many loci each with small mutational contributions, and that the effects of mutations are not too skewed.

Recalling that the moment generating function for a sum of independent random variables is the product of their generating functions, the moment generating function for a sample of size two in a panmictic population for a trait controlled by L loci is

$$\varphi(k_1, k_2) = \left(\frac{1}{1 - \frac{\theta}{2}(\psi(k_1) + \psi(k_2) - 2)} \right)^L. \quad (71)$$

If we rewrite this as

$$\left(1 + \frac{\frac{\theta}{2}L(\psi(k_1) + \psi(k_2) - 2)}{L - \frac{\theta}{2}L(\psi(k_1) + \psi(k_2) - 2)} \right)^L,$$

it's simple to show that in the limit as $L \rightarrow \infty$ we get

$$\exp \left(\frac{\theta}{2} \left(\frac{k_1^2}{2} \sigma^2 + \frac{k_2^2}{2} \sigma^2 \right) \right). \quad (72)$$

We can recognize this as the generating function of a multivariate normal distribution, meaning that

$$(Y_1, Y_2) \sim \mathcal{N} \left(\mathbf{0}, \begin{pmatrix} \frac{\theta}{2} \sigma^2 & 0 \\ 0 & \frac{\theta}{2} \sigma^2 \end{pmatrix} \right). \quad (73)$$

We next extend this logic to a sample of size three to see how shared branches induce correlations. The moment generating function for this case is

$$\begin{aligned} \varphi(k_1, k_2, k_3) = & \left(\frac{1}{3 - \frac{\theta}{2}(\psi(k_1) + \psi(k_2) + \psi(k_3) - 3)} \right. \\ & \times \left[\frac{1}{1 - \frac{\theta}{2}(\psi(k_1 + k_2) + \psi(k_3) - 2)} \right. \\ & + \frac{1}{1 - \frac{\theta}{2}(\psi(k_1 + k_3) + \psi(k_2) - 2)} \\ & \left. \left. + \frac{1}{1 - \frac{\theta}{2}(\psi(k_2 + k_3) + \psi(k_1) - 2)} \right] \right)^L \end{aligned} \quad (74)$$

We can already see that this is a product of two terms, and we can likely guess that it represents a convolution of two normal distributions. Rearranging this equation and simplifying notation a

bit gives

$$\begin{aligned} \varphi(k_1, k_2, k_3) = & \left(\frac{1}{1 - \frac{1}{3}\frac{\theta}{2}(\psi_1 + \psi_2 + \psi_3 - 3)} \right. \\ & \times \left[1 + \frac{\frac{1}{3}\frac{\theta}{2}(\psi_{12} + \psi_3 - 2)}{1 - \frac{\theta}{2}(\psi_{12} + \psi_3 - 2)} \right. \\ & + \frac{\frac{1}{3}\frac{\theta}{2}(\psi_{13} + \psi_2 - 2)}{1 - \frac{\theta}{2}(\psi_{13} + \psi_2 - 2)} \\ & \left. \left. + \frac{\frac{1}{3}\frac{\theta}{2}(\psi_{23} + \psi_1 - 2)}{1 - \frac{\theta}{2}(\psi_{23} + \psi_1 - 2)} \right] \right)^L. \end{aligned} \quad (75)$$

In the limit as $L \rightarrow \infty$ we get

$$\begin{aligned} \varphi(k_1, k_2, k_3) = & \exp\left(\frac{1}{3}\frac{\theta}{2}\sigma^2(k_1^2 + k_2^2 + k_3^2)\right) \\ & \times \exp\left(\frac{1}{3}\frac{\theta}{2}\sigma^2(3k_1^2 + 3k_2^2 + 3k_3^2 + 2k_1k_2 + 2k_1k_3 + 2k_2k_3)\right). \end{aligned} \quad (76)$$

This is the moment generating function of multivariate normal with means 0, variance $\frac{\theta}{2}\sigma^2\frac{4}{3}$, and covariance $\frac{\theta}{2}\sigma^2\frac{1}{3}$. Of course, we can recognize that $\frac{4}{3}$ is the expected TMRCA, and $\frac{1}{3}$ is the expected shard branch length between two samples. Applying this principle more generally, in a sample of size n the expected TMRCA will be $2(1 - \frac{1}{n})$ and the expected amount of shared branch length will be

$$E[T_{shared}] = 2 \sum_{i=1}^n \left(\frac{1}{i-1} - \frac{1}{i} \right) \left(1 - \prod_{j=i+1}^n \left(1 - \frac{2}{j(j-1)} \right) \right) = 1 - \frac{2}{n}. \quad (77)$$

I just solved this with Mathematica, but it should be possible to find this simple result by closer inspection. We should also get the same result from the generating function. The simplicity here is hopefully inspiring to be able to obtain a nice result in the structured population case. Interestingly, these values quickly converge as the sample size grows

1.10 The distribution of individual differences

In any sample, the absolute phenotypic values will not be meaningful. It is differences between individuals sampled from various sub populations that are of interest to us. We are therefore interested in the joint distribution of $Y_i - Y_j$ for each pair of individuals in the sample. Using the infinitesimal model with a distribution of mutational effects centered at zero the differences should have a multivariate normal distribution with mean zero. What matters is the covariances of the differences.

$$Cov[Y_1 - Y_2, Y_3 - Y_4] = Cov[Y_1, Y_3] + Cov[Y_2, Y_4] - Cov[Y_1, Y_4] - Cov[Y_2, Y_3] \quad (78)$$

We know from previous results that $Cov[Y_1, Y_2]$ is proportional to the shared branch length of individuals 1 and 2 before the most recent common ancestor of the sample. We therefore have

$$Cov[Y_1, Y_2] \propto E[T_{MRC A}] - E[\tau_{1,2}], \quad (79)$$

and

$$\text{Cov}[Y_1 - Y_2, Y_3 - Y_4] \propto E[\tau_{1,4}] + E[\tau_{2,3}] - E[\tau_{1,3}] - E[\tau_{2,4}]. \quad (80)$$

Where $\tau_{1,2}$ is the expected coalescence time of individuals 1 and 2.

1.10.1 The sampling distribution of Q_{ST} under the infinitesimal model

The Q_{ST} statistic measures how much of the phenotypic variation in a structure population is partitioned between groups. Q_{ST} is defined as

$$Q_{ST} := \frac{V_{\text{between}}}{V_{\text{between}} + V_{\text{within}}}. \quad (81)$$

In the usual analysis of variance framework, the variance between K populations is $\frac{1}{K} \sum (\bar{Y}_i - \bar{Y})^2$ and the variance within groups is $\frac{1}{\sum N_k} \sum_i \sum_j (Y_{i,j} - \bar{Y}_i)^2$. To examine the distribution of Q_{ST} at the population level one can figure out what the distributions of V_{between} and V_{within} are over evolutionary realizations. Even if the full distribution of Q_{ST} does not have an analytic form, we can sample from the distributions of the variance to calculate tail probabilities.

Under the infinitesimal model, each $Y_{i,j}$ is normally distributed, so the means are therefore also normal. When the population size is large enough, $(Y_{i,j} - \bar{Y}_i)^2$ and $(Y_{j,l} - \bar{Y}_j)^2$ are nearly uncorrelated both within and between populations because each individual contributes little to the population means. V_{within} is therefore a sum of squared independent normal random variables with mean zero. With enough individuals in the whole population, the central limit theorem kicks in and the variance is order $1/\sum N_k$. The mean value of $E[(Y_{i,j} - \bar{Y}_i)^2]$ is $\sigma^2 E[\tau_{i,i}]$. We can thus treat V_{within} as a constant with value $\frac{\sum N_k E[\tau_{k,k}]}{\sum N_k}$. If c_k is the fraction of the total population in group k , then $V_{\text{within}} = \sum c_k E[\tau_{k,k}]$. This brings up the issue that the fraction that each group makes up of the total population is unlikely to be known. This makes thinking about Q_{ST} as a population parameter a bit tricky. I think it is best to think of it as an idealized sample value when a large and equal number of samples is taken from each population.

In the variance between groups, the $(\bar{Y}_i - \bar{Y})^2$ terms are correlated. The sum of these will have a generalized chi-square distribution. We could sample from this distribution by simulating the $\bar{Y}_i - \bar{Y}$ from a multivariate normal distribution and adding the squared values. The covariance matrix for this distribution is

$$\text{Cov}[\bar{Y}_i - \bar{Y}, \bar{Y}_j - \bar{Y}] = \sigma^2 (E[\tau_{i,\cdot}] + E[\tau_{j,\cdot}] - E[\tau_{\cdot,\cdot}] - E[\tau_{i,j}]). \quad (82)$$

2 Dominance

2.1 The individual variance

When modeling dominance each individual receives two copies of alleles from the population. These may contain the same or different alleles. An individual's genotype is then made up of contributions from L loci.

$$Y = \sum_{l=1}^L f(Y_{l,1}, Y_{l,2}). \quad (83)$$

The function f is a function that describes the "dominance" relationship for the trait. It is natural to set $f(0,0) = 0$ so that there is no effect when an individual receives no mutations at a locus. If we assume that each locus has an independent genealogy, then the variance of Y can

be computed by calculating $\text{Var}[f(Y_{l,1}, Y_{l,2})]$. One way to do this is by using the law of total variance and conditioning on the mutational configuration at the locus.

$$\text{E}[\text{Var}(f(Y_{l,1}, Y_{l,2})|\text{mutation})] + \text{Var}(\text{E}[f(Y_{l,1}, Y_{l,2})|\text{mutation}]).$$

The second term here will be zero if the mutational distribution and transformation f are both symmetric. We'll then assume that at most one mutation occurs per locus genealogy. This means it is only necessary to calculate $\text{Var}(f(Y_{l,1}, Y_{l,2})|1 \text{ mutation})$ and $\text{Var}(f(Y_{l,1}, Y_{l,2})|2 \text{ mutations})$.

To proceed one will need to assume a specific form for f . A very simple function is

$$f(Y_{l,1}, Y_{l,2}) = \begin{cases} bY_{l,1}, & \text{if } Y_{l,2} = 1 \\ 2Y_{l,1}, & \text{if } Y_{l,2} \neq 1. \end{cases} \quad (84)$$

In this model every mutation has the same degree of dominance regardless of how big of an effect it has on the trait. More complicated models are possible, but this one is good so far for building intuition. Under this model,

$$\text{Var}(f(Y_{l,1}, Y_{l,2})|1 \text{ mutation}) = b^2 m_2,$$

and

$$\text{Var}(f(Y_{l,1}, Y_{l,2})|2 \text{ mutations}) = 4m_2.$$

The expected variance at a locus is then

$$\text{Var}(f(Y_{l,1}, Y_{l,2})|1 \text{ mutation})\P(1 \text{ mutation}) + \text{Var}(f(Y_{l,1}, Y_{l,2})|2 \text{ mutations})\P(2 \text{ mutations}).$$

Which evaluates overall to

$$2\frac{\theta}{2}\tau_{2,2}b^2m_2 + \frac{\theta}{2}(T_{MRCA} - \tau_{2,2})4m_2. \quad (85)$$

We only need to take the expectation over genealogies at each locus to finish the derivation

2.2 The population variance

The variance in trait values between individuals determines the expected variance in the population. This variance can be written as

$$\text{Var}\left(\sum_{l=1}^L f(Y_{1,l,1}, Y_{1,l,2}) - \sum_{l=1}^L f(Y_{2,l,1}, Y_{2,l,2})\right).$$

The important term to calculate is $\text{Cov}(f(Y_{1,l,1}, Y_{1,l,2}), f(Y_{2,l,1}, Y_{2,l,2})) = \text{E}[f(Y_{1,l,1}, Y_{1,l,2}), f(Y_{2,l,1}, Y_{2,l,2})]$. We can use the law of total expectation to calculate this again conditioning on different mutational configurations. The configurations that lead to a nonzero expectation if one mutation occurs is present in each individual, three mutations occur, or four mutations occur. There are four ways for the two individuals to each have one mutant copy. When this occurs the expected product will be $b^2 m_w$. There are four ways for a mutation to be present in three copies and when this occurs the expected product will be $2bm_2$. There is one way to have a mutation present in all four copies and the expected product is $4m_2$.

We can then use the probability of each type of mutational configuration occurring to calculate the covariance.

3 Epistasis

3.1 Nonlinear transformation of trait values

For now this section is exploratory, so the notation will likely differ from the rest of the notes.

One simple form of epistasis is a nonlinear transformation of the individual trait values. If X_i are the genetic values contributed by each of L loci, then the trait value of an individual would be

$$Y = f\left(\sum_{i=1}^L X_i\right).$$

Under the infinitesimal model the $\sum X_i$ are normally distributed. As a nonlinear transformation of a normal variable, Y will not be normally distributed. However, we can still investigate its first two moments to see the effects of epistasis. If f is very complicated then this could be a very difficult thing to do. We can start by taking a second-order Taylor expansion of f . Expanding f around zero gives

$$Y \approx f(0) + f'(0) \sum_{i=1}^L X_i + \frac{1}{2} f''(0) \left(\sum_{i=1}^L X_i\right)^2.$$

This assumes that the curvature is the same in each direction, which is not a crazy assumption. Let $f(0) = 0$, $f'(0) = 1$, and $f''(0) = \epsilon$. When deriving stuff under the assumption that $\sum X_i$ is normal, let $X = \sum X_i$.

The first thing to investigate is the effect of weak epistasis on the variance. This is

$$\text{Var}[X + \epsilon X^2] = \text{Var}[X] + \epsilon \text{Var}[X^2] + 2\epsilon \text{Cov}[X, X^2].$$

Using standard results on the normal distribution and previously derived results, $\text{Var}[X] = \mu \text{ET}_{MRCA}$, $\text{Var}[X^2] = 2\sigma^4 \text{ET}_{MRCA}^2 + 4\sigma^2 \mu^2 \text{ET}_{MRCA}^3$, and $\text{Cov}[X, X^2] = 2\mu\sigma^2 \text{ET}_{MRCA}$. Putting this together and setting $\mu = 0$ gives the variance

$$\text{Var}[X + \epsilon X^2] = \sigma^2 \text{ET}_{MRCA} + 2\epsilon\sigma^2 \text{E}[T_{MRCA}]^2. \quad (86)$$

The corresponding expectation is

$$\text{E}[X + \epsilon X^2] = \mu \text{ET}_{MRCA} + \epsilon \sigma^2 \text{ET}_{MRCA}. \quad (87)$$

The covariance can be derived in a similar manner.

$$\text{Cov}[X_1 + \epsilon X_1^2, X_2 + \epsilon X_2^2] = \text{Cov}[X_1, X_2] + \epsilon \text{Cov}[X_1, X_2^2] + \epsilon \text{Cov}[X_1^2, X_2] + \epsilon^2 \text{Cov}[X_1^2, X_2^2]. \quad (88)$$

If $\mu = 0$ the second and third term are zero and the fourth term is

$$2\sigma^4 (\text{ET}_{MRCA} - \tau_{1,2})^2.$$

This ultimately gives

$$\text{Cov}[X_1 + \epsilon X_1^2, X_2 + \epsilon X_2^2] = \sigma^2 (\text{ET}_{MRCA} - \tau_{1,2}) + 2\epsilon^2 \sigma^4 (\text{ET}_{MRCA} - \tau_{1,2})^2. \quad (89)$$

Additionally, the variance in the difference in trait values between sampled individuals is a good measure of the amount of variation in the population. Indeed, as long as the individuals are exchangeable this is proportional to the expected variance in a large population.

$$\text{Var}[X_1 - \epsilon X_1^2 - X_2 - \epsilon X_2^2] = \text{Var}[X_1 + \epsilon X_1^2] + \text{Var}[X_2 + \epsilon X_2^2] - 2\text{Cov}[X_1 + \epsilon X_1^2, X_2 + \epsilon X_2^2].$$

Again, if $\mu = 0$ we can use previous results to get

$$\text{Var}[X_1 - \epsilon X_1^2 - X_2 - \epsilon X_2^2] = 2\sigma^2\tau_{1,2} + 4\sigma^4\epsilon \left(\text{E}[T_{MRC A}]^2 - \epsilon\sigma^2(\text{E}T_{MRC A} - \tau_{1,2})^2 \right).$$

What is interesting about this is that we can see when the expected population variance will be greater than that without epistasis. This is the case when

$$\frac{\text{E}[T_{MRC A}]^2}{\text{E}[T_{MRC A} - \tau_{1,2}]^2} > \epsilon\sigma^2 \quad (90)$$

When $\epsilon < 0$ the variance is always less than under the additive case.

4 Selection

4.1 Selection on a polygenic character

We have investigated the fact that, when the number of loci expected to experience mutations affecting a trait is low, the kurtosis is higher than that for a normal distribution. This kurtosis is over evolutionary realizations. If we were to replay evolution then individuals would have more extreme values relative to the variance of the distribution than under a normal distribution. It is worth asking how this affects selection. Even though the models so far have been strictly neutral, we can imagine selection acting on the population and producing a change in the mean and variance of trait values. The classical theory of quantitative genetics assumes a normal distribution of breeding values in the population. In one of a series of papers, TURELLI and BARTON (1990) extended the theory of selection on a polygenic character to a more realistic model of multilocus population genetics. One conclusion of this work is that selection can cause deviations from normality in higher order moments of the distribution of breeding values, affecting the progress of selection.

A central result of TURELLI and BARTON (1990) is that the change in the mean phenotype due to one generation of selection is

$$\Delta \bar{Z} = V_g L_1 + M_{3,g} L_2 + \gamma_4 V_g^2 L_3 + (M_{5,g} - 4M_{3,g} V_g) L_4 + \dots \quad (91)$$

Here, Z refers to a phenotypic value which has an environmental component as opposed to the breeding value Y which is due entirely to genetics. V_g is the variance in breeding values in the population and γ_4 is the excess kurtosis above a normal distribution. The terms $M_{i,g}$ are the i^{th} central moments of the breeding value distribution in the population. The terms L_i describe the effect of selection on the breeding values. These are the selection gradients in terms of the genotypic moments.

$$L_i = \frac{\partial \ln(\bar{w})}{\partial M_{i,g}} \quad (92)$$

and

$$L_1 = \frac{\partial \ln(\bar{w})}{\partial \bar{Y}}. \quad (93)$$

What we can tell immediately from these equations is that the importance of higher order moments of the breeding value distribution depends on the specific fitness function. For instances, the effect of the skew of the breeding value distribution $M_{3,g}$ depends on the effect on the mean

fitness of changing the variance of the breeding values $\frac{\partial \ln(\bar{w})}{V_g}$. Each of the L_i terms correspond to the effects on mean fitness of changing one moment while holding the others constant. Therefore, whether or not the excess kurtosis that arises due to a sparse trait architecture has a meaningful effect on the response to selection depends on the precise shape of that selection. TURELLI and BARTON (1990) use trick to calculate the L_i . By taking a Taylor series expansion of $w_g(Y)$ they get

$$\bar{w} = w_g(\bar{Z}) + \sum_{i=2}^{\infty} \frac{M_{i,g} w_g^{(i)}(\bar{Z})}{i!}. \quad (94)$$

Differentiating this gives

$$L_1 = \frac{w_g^{(1)}(\bar{Z})}{\bar{w}} + \sum_{i=2}^{\infty} \frac{M_{i,g} w_g^{(i+1)}(\bar{Z})}{i! \bar{w}}, \quad (95)$$

and

$$L_i = \frac{w_g^{(i)}(\bar{Z})}{i! \bar{w}}. \quad (96)$$

What this shows is that it is sufficient when calculating L_i to calculate the i^{th} derivative of w_g and evaluate this at the population mean.

4.2 Exponential selection

One simple fitness function is exponential directional selection:

$$w(z) = e^{sz}. \quad (97)$$

This fitness function has a number of nice properties. Fitness is multiplicative across loci so it does not lead to a build up of linkage disequilibrium. Additionally, the ratio of fitnesses between two phenotypes does not depend on the reference from which one measures them. TURELLI and BARTON (1990) investigate exponential directional selection. They calculate

$$\bar{w} = e^{sz} \exp\left(\frac{s^2 V_e}{2}\right) \left(1 + \sum_{i=2}^{\infty} \frac{s^i M_{i,g}}{i!}\right), \quad (98)$$

$$L_1 = s, \quad L_2 \approx s^2, \quad \text{and} \quad L_k \approx 0 \quad \text{for} \quad k \geq 3. \quad (99)$$

The approximations ignore terms of order s^2 . Thanks to (91) we can see that only the variance and skew of the breeding values will affect the response to selection in this case.

4.3 Cubic selection

The population kurtosis may become more important if we want to consider selection where the fitness function is cubic on the breeding values. Such a fitness function would have the form

$$W_g(Y) = b_0 + b_3(Y - \bar{Y})^3. \quad (100)$$

This fitness function now models selection on differences from the current mean breeding value in the population. This is not the most realistic scenario because this mean will have diverged

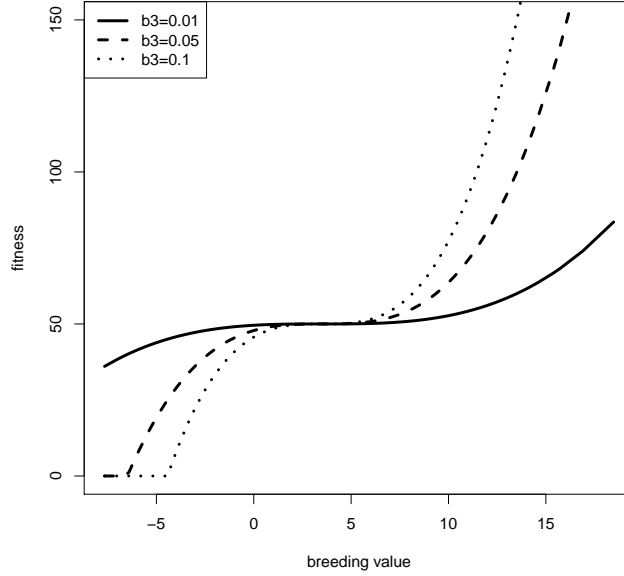


Figure 2: Cubic fitness functions with $b_0 = 50$ applied to a random population with 2000 individuals, 100 loci, and a standard deviation of mutational effects of one.

since the most recent common ancestor of the population in the absence of stabilizing selection. The reference point which selection sees as the center is therefore somewhat arbitrary. There's no good reason for this to be the population mean, but it's also not totally unreasonable. Another thing we need to consider about the cubic selection function is that it can give negative values for fitness, which is something we clearly don't want. We therefore must assume that the degree of cubic selection, b_3 is not too large relative to fitness value at the population mean, b_0 . Figure 2 shows a few examples of cubic fitness functions applied to a simulated population with mean breeding value around four. Selection gets very steep as b_3 increases and one ends up with negative fitness values that are then set to zero

Note that we're now considering selection directly on breeding values rather than starting with selection on phenotypic values and deriving selection on breeding values assuming a Gaussian distribution for environmental effects. We might justify this by saying that this should capture the effect of that portion of the overall fitness function which has a cubic shape with respect to the breeding values.

Applying (92), (94), and (94) to (100) we get

$$\bar{w} = b_0 + M_{3,g}b_3, \quad (101)$$

$$L_1 = \frac{3V_g b_3}{\bar{w}}, \quad (102)$$

$$L_2 = 0, \quad (103)$$

$$L_3 = \frac{b_3}{\bar{w}}. \quad (104)$$

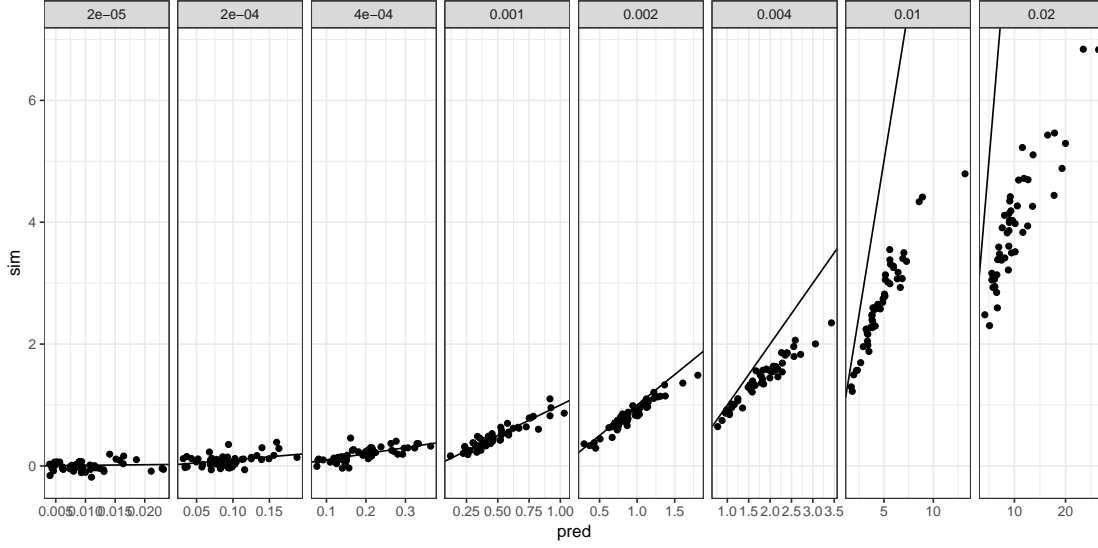


Figure 3: Simulation results comparing the response to varying strengths of cubic selection to the expected response under weak selection theory given by (105). Simulation realizations for a given β have different responses to selection because of differences in the fourth moment of breeding values arising by chance. Lines show the one-to-one relationship for comparison.

After combining these all into (91) we get

$$\Delta \bar{Z} = \frac{3V_g^2 b_3 + (M_{4,g} - 3V_g^2) b_3}{b_0 + M_{3,g} b_3} = \frac{M_{4,g} b_3 / b_0}{1 + M_{3,g} b_3 / b_0} = \frac{M_{4,g} \beta}{1 + M_{3,g} \beta}. \quad (105)$$

We can see that the response to cubic selection depends only on β , the ratio of the steepness of cubic selection to the baseline fitness, as well as the third and fourth moments of the distribution of breeding values. In general the response to cubic selection is linear with respect to the fourth moment of the distribution of breeding values and hence the kurtosis.

Since the theory of TURELLI and BARTON (1990) is for weak selection, we must investigate the range of selection strengths for which (100) predicts the average response to selection. To do this, for a range of β values I simulated 50 populations with 2000 haploid individuals each and a mutational standard deviation of one. I drew individuals for the next generation according to fitness values from (100) and compared the change in phenotype to that predicted by (105). The results of this experiment are shown in Figure 4.3. The weak selection approximations of TURELLI and BARTON (1990) begin to break down for β greater than 0.001, which as Figure 2 shows, corresponds to quite strong selection.

References

- KIMURA, M., 1965 A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences* **54**: 731–736.
- KINGMAN, J., 1978 A Simple Model for the Balance between Selection and Mutation. *Journal of Applied Probability* **15**: 1–12.
- LOHSE, K., R. J. HARRISON, and N. H. BARTON, 2011 A general method for calculating likelihoods under the coalescent process. *Genetics* **189**: 977–87.
- SCHRAIBER, J. G., and M. J. LANDIS, 2015 Sensitivity of quantitative traits to mutational effects and number of loci. *Theoretical population biology* **102**: 85–93.
- TURELLI, M., and N. H. BARTON, 1990 Dynamics of polygenic characters under selection. *Theoretical Population Biology* **38**: 1–57.

5 Supplement

5.1 Additional steps for some derivations

In the derivation of (34), how the second term is arrived at is not immediately obvious. To see where this comes from, consider the two types of terms within (??) that will contribute $k_a^3 k_b$ from pairs of branches. These are those pairs where both contain a and b , and those pairs where one contains a and b and the other contains a only. Pairs with the same branch repeated twice have multiplicity $L(L-1)/2$ and pairs with two different branches have multiplicity $L(L-1)$. Branch pairs where both contain a and b have $4k_a^3 k_b$ because this term can be made four times from $(k_a + k_b)^4$, and branch pairs where one contains a and the other a and b have $2k_a^3 k_b$ because this term can be made two ways from $k_a^2(k_a + k_b)^2$. These two sets of pairs are

$$\frac{L(L-1)}{2} \left(\frac{\theta m_2}{2} \right)^2 2k_a^2 \times 2k_a k_b \left(\sum_{\omega: a, b \in \omega} E[t_\omega] \right)^2$$

and

$$L(L-1) \left(\frac{\theta m_2}{2} \right)^2 k_a^2 \times 2k_a k_b \left(\sum_{\omega: a, b \in \omega} E[t_\omega] \right) \left(\sum_{\omega: a/b \in \omega} E[t_\omega] \right).$$

These two can clearly be added together to yield

$$L(L-1) \left(\frac{\theta m_2}{2} \right)^2 k_a^2 \times 2k_a k_b \left(\sum_{\omega: a, b \in \omega} E[t_\omega] \right) \left(\sum_{\omega: a \in \omega} E[t_\omega] \right).$$

The derivation of (35) is even more challenging because more potential branch pairs need to be considered between the three descendants included in the moment. As before pairs containing the same branch twice have a multiplicity of $L(L-1)/2$ while pairs containing two different branches have a multiplicity of $L(L-1)$. Depending on what individuals are on each branch $k_a^2 k_b k_c$ will also have a different coefficient. I'll consider each possible type of pair in sequence then add them together at the end.

a , b , and c are present on all branches

Since all individuals are present on both branches, the coefficient of $k_a^2 k_b k_c$ is multinomial (12). The term for this set of pairs is then

$$\frac{L(L-1)}{2} 12 k_a^2 k_b k_c \left(\sum_{\omega: a, b, c \in \omega} E[t_\omega] \right)^2.$$

Only a and b are on one branch while a , b and c are on the other

In this case $k_a^2 k_b k_c$ has coefficient 6. Since these are non-overlapping sets of branches we get

$$L(L-1) 6 k_a^2 k_b k_c \left(\sum_{\omega: a, b, c \in \omega} E[t_\omega] \right) \left(\sum_{\omega: a, b, c \in \omega} E[t_\omega] \right).$$

Only a and c are on one branch while a , b and c are on the other

This is the same as above.

$$L(L-1) 6 k_a^2 k_b k_c \left(\sum_{\omega: a, c, b \in \omega} E[t_\omega] \right) \left(\sum_{\omega: a, b, c \in \omega} E[t_\omega] \right).$$

Only a is present on one branch while a , b and c are on the other

In this case $k_a^2 k_b k_c$ has coefficient 2. These branches contain non-overlapping sets of descendants again so we can write

$$L(L-1) 2 k_a^2 k_b k_c \left(\sum_{\omega: a, b, c \in \omega} E[t_\omega] \right) \left(\sum_{\omega: a, b, c \in \omega} E[t_\omega] \right).$$

a is present on one branch while only b and c are present on the other

In this case, no matter what other descendants join a on its branch, the coefficient of $k_a^2 k_b k_c$ will be 2. These branches are also non-overlapping, so we get

$$L(L-1) 2 k_a^2 k_b k_c \left(\sum_{\omega: a \in \omega} E[t_\omega] \right) \left(\sum_{\omega: b, c, a \in \omega} E[t_\omega] \right).$$

Only a and b are present on one branch while only a and c are present on the other.

In this final case the coefficient of $k_a^2 k_b k_c$ is 4 because we are taking this term from $(k_a + k_b)^2 (k_a + k_c)^2$. These are again different branches always so

$$4L(L-1) k_a^2 k_b k_c \left(\sum_{\omega: a, b, c \in \omega} E[t_\omega] \right) \left(\sum_{\omega: a, c, b \in \omega} E[t_\omega] \right).$$

This can be simplified with a small amount of branch arithmetic. Being very lazy about notation this is

$$\begin{aligned}
& 6\tau_{a+b+c}^2 + 6\tau_{a+b/c}\tau_{a+b+c} + 6\tau_{a+c/b}\tau_{a+b+c} + 2\tau_{a/b+c}\tau_{a+b+c} + 2\tau_a\tau_{b+c/a} + 4\tau_{a+b/c}\tau_{a+c/b} \\
& = 4\tau_{a+b+c}^2 + 4\tau_{a+b/c}\tau_{a+b+c} + 4\tau_{a+c/b}\tau_{a+b+c} + 2\tau_a\tau_{a+b+c} + 2\tau_a\tau_{a+b} - 2\tau_a\tau_{a+b+c} + 4\tau_{a+b/c}\tau_{a+c/b} \\
& \quad = 2\tau_a\tau_{a+b} + 4\tau_{a+b+c}^2 + 8\tau_{a+b/c}\tau_{a+b+c} + 4\tau_{a+b/c}^2 \\
& \quad = 2\tau_a\tau_{a+b} + 4(\tau_{a+b+c} + \tau_{a+b/c})^2 \\
& \quad = 2\tau_a\tau_{a+b} + 4\tau_{a+b}^2.
\end{aligned}$$

This then gives us the same result as (35).