Lecture 21
The infinitesimal model and its extensions

Bruce Walsh lecture notes
Synbreed course
version 9 July 2013

Selection (and drift) compromise predictions of selection response by changing allele frequencies and generating disequilibrium, changing Var(*A*) and hence h^2.

Predicting changes in allele frequencies are especially problematic, requiring very intimate knowledge of the underlying genetical details, such as the effects and frequencies of all alleles. As these are essentially unobservable for all but the most trivial cases, these are often referred to as the microscopic (unobservable) parameters of the system.

Ideally, we would like to have macroscopic-based (observable) predictors of response, based on easily-measured quantities, such as genetic variances.

Additive variance with LD:

Additive variance is the variance of the sum of allelic effects,

$$\sigma^2\left(\sum_{k=1}^{n}\left(a_1^{(k)}+a_2^{(k)}\right)\right)=2\sum_{k=1}^{n}\sigma^2\left(a^{(k)}\right)+4\sum_{k<j}^{n}\sigma\left(a^{(j)},a^{(k)}\right)$$

$$=2\sum_{k=1}^{n}C_{kk}+4\sum_{k<j}^{n}C_{jk}$$

$$\sigma_A^2=\sigma_a^2+d$$

Genic variance: value of Var(A) in the absence of disequilibrium function of allele frequencies

Disequilibrium contribution (d). Requires covariances between allelic effects at different loci = LD

Genetic vs. genic variance

- Recall that the genic variance, namely that value of the additive variance for the current set of allele frequencies in the absence of LD is entirely a function of allele frequencies
  - Under the infinitesimal model (in an infinite population size), the genic variance is unchanged by selection
  - Conversely, the change in disequilibrium (d) is entirely predictable as well (Bulmer equation)
  - Likewise, under strict drift, the genic variance (assuming additivity) declines in a very predicable manner.

Under the infinitesimal model, allele frequencies are unchanged by selection, and thus the genic variance σ^2_a is assumed constant.

Large changes in the mean occur by summing infinitesimal allele frequency changes over a large number of loci

Suppose n loci, allele freq p, effect a. Mean = 2npa

$\sigma^2_A = \sigma^2_a = 2na^2p(1-p)$, so a must scale as $n^{-1/2}$ for variance to remain bounded for large n.

$$\Delta\mu = 2na\Delta p, \quad \text{thus} \quad \Delta p = \Delta\mu/(2na)$$

Since a is of order $n^{-1/2}$, Δp is of order $1/(n \cdot n^{-1/2}) = n^{-1/2}$

$$\begin{aligned}\Delta\sigma_a^2 &= 2na^2p'(1-p') - 2na^2p(1-p)\\ &= 2na^2\Delta p(1-2p-\Delta p)\\ &\approx a\,(1-2p)\,\Delta\mu\end{aligned}$$

Since a is order $n^{-1/2}$, changes in variance of order of $\Delta\mu/n^{1/2}$

Bottom line: When the number of loci is large, and each have roughly equal effect, can get very larges changes in the mean with no appreciable change in the allele frequencies or (more importantly) the genic variance

Gaussian continuum-of-alleles (COA) models

Kimura found that if new mutations have small effects relative to the existing variation at the locus, then the distribution of effects (in an infinite population) converges to a normal.

COA models make the assumption that the distribution of breeding values *at each locus* is gaussian (and jointly multivariate normal over a vector of loci), which can only be strictly correct for an infinite number of alleles / locus

This assumption of a Gaussian distribution of effects at each locus is much more restrictive than the assumption that the distribution of the *total* genotypic value is normal.

While the distribution of total genotypic values is gaussian under continuum-of-alleles model, the central limit theorem allows the sum of non-normal distributions across loci to converge to a Gaussian.

Thus, COA models are a very restrictive subset of all possible models that can lead to the infinitesimal.

The advantage of COA models is that we can assume a finite number of loci and hence partly accommodate allele frequency change.

Gaussian COA models attempt to bridge the infinitesimal on one hand with models allowing allele frequency change on the other (i.e., models where the genic variance changes.

Under *COA* assumptions, consider pure drift. While selection does not change allele freq, drift does, with a simple expression for the expected change in genic variance being

$$\Delta\,\sigma_a^2(t) = -\frac{\sigma_a^2(t)}{2N_e}$$

$$\Delta\,d(t) = -\frac{1}{2}\left[\left(1+\frac{1}{N_e}\right)d(t) + \left(1-\frac{1}{N_e}\right)\kappa\,h^2(t)\,\sigma_A^2(t)\right]$$

Drift also generates a small amount of LD

More generally, suppose there are n loci of equal effect as well as drift (finite N_e)

$$\Delta\,\sigma_a^2(t) = -\left[\frac{\sigma_a^2(t)}{2N_e} + \left(1-\frac{1}{N_e}\right)\frac{\kappa\,h^2(t)\,\sigma_A^2(t)}{2n}\right]$$

$$\Delta\,d(t) = -\frac{1}{2}\left[\left(1+\frac{1}{N_e}\right)d(t) + \left(1-\frac{1}{n}\right)\left(1-\frac{1}{N_e}\right)\kappa\,h^2(t)\,\sigma_A^2(t)\right]$$

Change in the genic variance and d scale as 1/n

The effect of drift on d is fairly small and can usually be ignored

What does response look like under these Models?

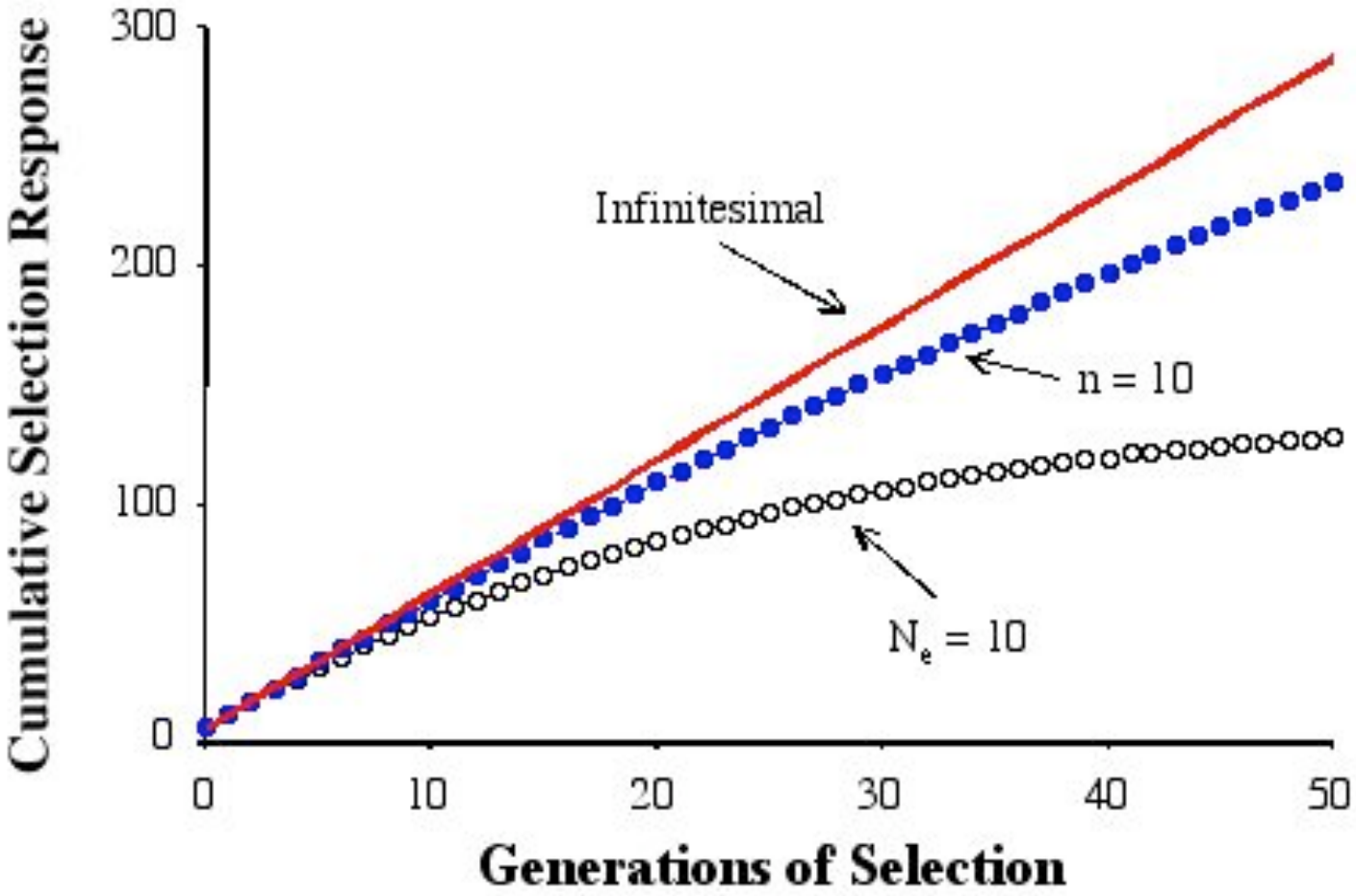
Again, the additive variation in generation t is

$$\sigma^2_A(t) = \sigma^2_a(t) + d(t)$$

$$\sigma^2_z(t) = \sigma^2_A(t) + \sigma^2_E$$

$$h^2(t) = \sigma^2_A(t) / \sigma^2_z(t)$$

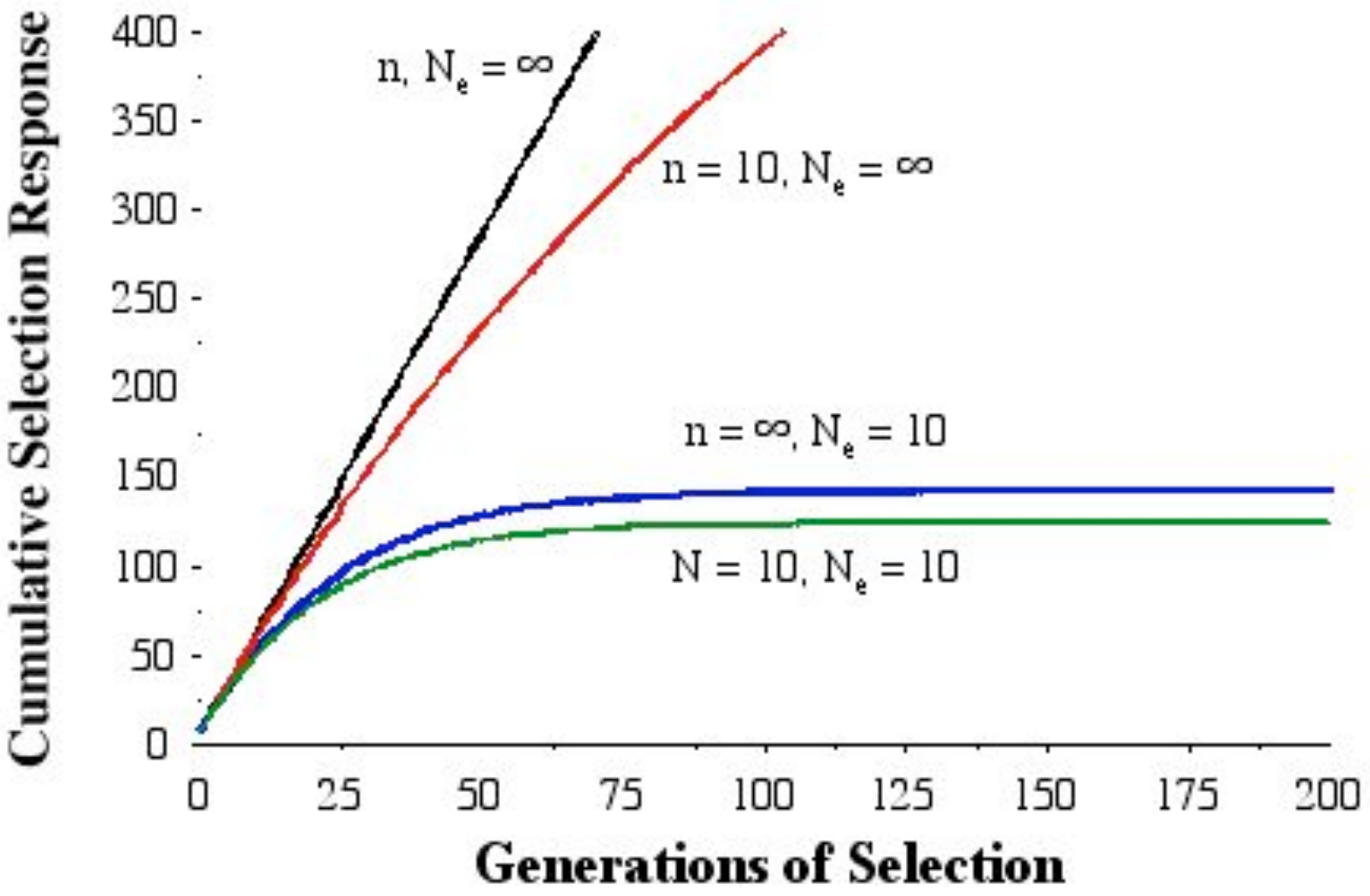
$$R(t) = h^2(t)S(t)$$



(i) Infinitesimal Ne = n = ∞

(ii) Ne = 10, n = ∞

(iii) Ne = ∞, n = 10


Cumulative Selection Response
400
350
300
250
200
150
100
50
0
n, $N_e = \infty$
n = 10, $N_e = \infty$
n = ∞, N_e = 10
N = 10, N_e = 10
0
25
50
75
100
125
150
175
200
Generations of Selection

When loci vary in effects, an important concept is the effective number of loci, n_e

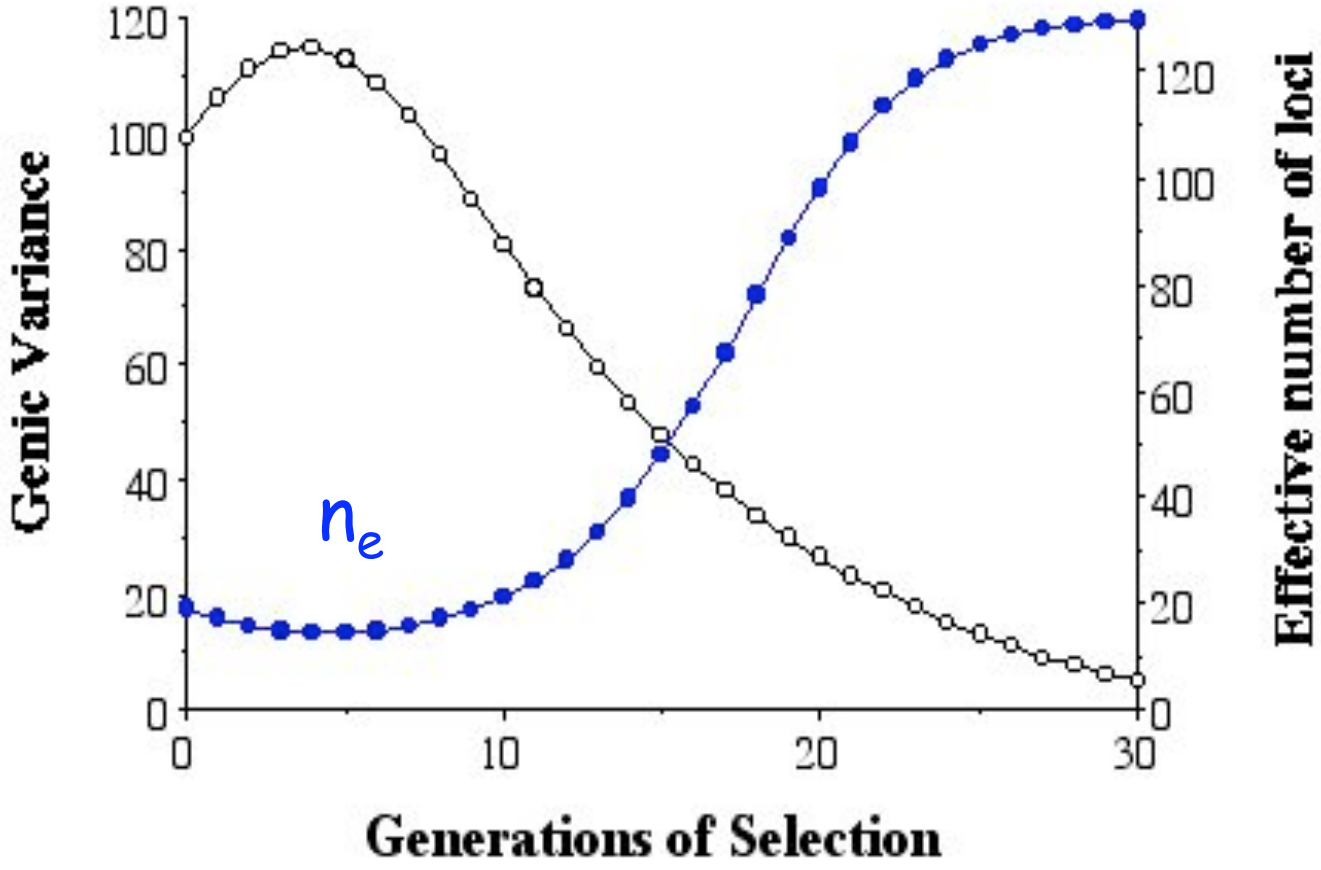
Replace n by n_e, $$n_e = \frac{n}{1 + cv^2}$$

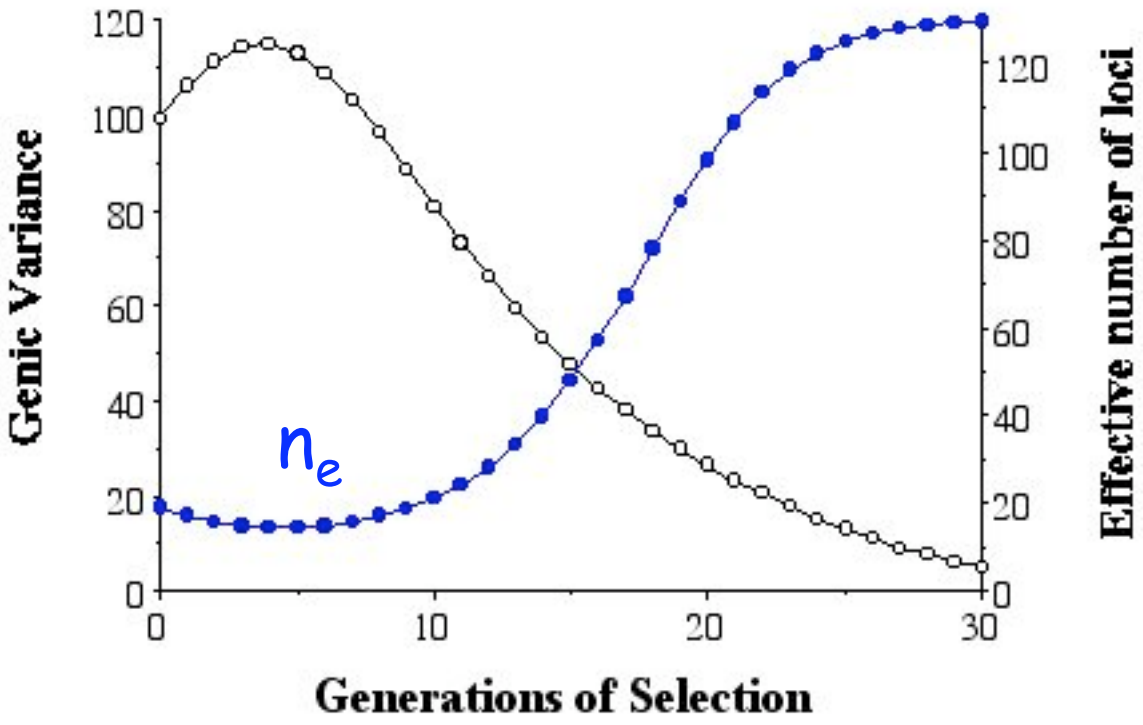
cv is the coefficient of variation in the genic variance contributed by each locus

$$cv = \sigma\left(\sigma^2_{ai}\right) / E[\,\sigma^2_{ai}\,]$$

Here σ^2_{ai} is the genic variance contributed by locus i.

Example 24.3. Consider an additive model with both major and minor loci. There are five major loci, each with frequency $p = 0.25$ and effect $a = 5.16$, and 125 minor loci, each with $p = 0.5$ and $a = 0.89$. The resulting initial genic variance is $\sigma_a^2 = 100$ and we assume a initial heritability of $h^2 = 0.5$. Finally, we assume truncation selection with the uppermost 20% saved (further details for this model are given in Example 26.2). We ignore any effects of disequilibrium, focusing on how the genic variance (open circles) and the effective number of loci n_e (filled circles) change over time due simply to allele frequency changes.





While there are 130 loci in this system, initially the effective number is around 20, due to the large coefficient of variance in the locus-specific genic variances. As we start selection, the additive variance initially increases, as the major alleles increase their frequencies toward 0.5 (where they have maximal additive variance). Such an increase in variance is not predicted by COA models. Notice that the effective number of loci further decreases during this increase in variance, as the coefficient in variation for the locus-specific variance increases as the genic variances at each major locus increases. As these major loci become fixed, the total genic variance decreases, while the effective number of loci increases, reflecting a decrease in the coefficient of variation.

In class project

Assume an initial $h^2 = 0.3$, $d(0) = 0$, phenotypic variance of 100, and the upper 2% selection ($\kappa = 0.89$, $i = 2.1$) Compute the cumulative response over 10 generations for

- Ne = 1000, n = 5
- Ne = 100, n = 100
- Ne = n = ∞

$$\Delta \sigma_a^2(t) = -\left[\frac{\sigma_a^2(t)}{2N_e} + \left(1 - \frac{1}{N_e}\right) \frac{\kappa\, h^2(t)\, \sigma_A^2(t)}{2n} \right]$$

$$\Delta d(t) = -\frac{1}{2}\left[\left(1 + \frac{1}{N_e}\right) d(t) + \left(1 - \frac{1}{n}\right)\left(1 - \frac{1}{N_e}\right) \kappa\, h^2(t)\, \sigma_A^2(t) \right]$$

Barton and Turelli: Nongaussian distributions of genotypic effects

Turelli and Barton examined selection response under general models

Take home message: Models do not close

- Genotypic variance & skew needed to predict response in mean
- Skew & Kurotsis needed to predict response in Variance, and so on

$$\boldsymbol{\Delta}_{\mu_G} \simeq \mathbf{M}\,\nabla \ln \overline{w}$$

$$\boldsymbol{\Delta}_{\mu_G} = \begin{bmatrix} \Delta\mu_{1,G} \\ \Delta\mu_{2,G} \\ \Delta\mu_{3,G} \\ \vdots \end{bmatrix}, \qquad \nabla \ln \overline{w} = \begin{bmatrix} \dfrac{\partial \ln \overline{w}}{\partial \mu_{1,z}} \\ \dfrac{\partial \ln \overline{w}}{\partial \mu_{2,z}} \\ \dfrac{\partial \ln \overline{w}}{\partial \mu_{3,z}} \\ \vdots \end{bmatrix}$$

$$\mathbf{M} = 2\sum_i \begin{bmatrix} \mu_{2,i} & \mu_{3,i} & (\mu_{4,i} - 3\mu_{2,i}^2) & \cdots \\ \mu_{3,i} & (\mu_{4,i} - \mu_{2,i}^2) & (\mu_{5,i} - 4\mu_{3,i}\,\mu_{2,i}) & \cdots \\ (\mu_{4,i} - 3\mu_{2,i}^2) & (\mu_{5,i} - 4\mu_{3,i}\,\mu_{2,i}) & (\mu_{6,i} - \mu_{3,i}^2 - 6\mu_{2,i}\,\mu_{4,i} + 9\mu_{2,i}^2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$\Delta\mu_G \simeq h^2 S + \left(\frac{\delta(\sigma_z^2) + S^2}{2\sigma_z^4}\right) M_3$$

$$\Delta\sigma_A^2 \simeq \frac{S}{\sigma_z^2} M_3 + \left(\frac{\delta(\sigma_z^2) + S^2}{\sigma_z^4}\right) \sum_i^n \left(\mu_{4,i} - \mu_{2,i}^2\right)$$

$$\Delta\mu_{3,G} \simeq \frac{S}{\sigma_z^2} M_4 + \left(\frac{\delta(\sigma_z^2) + S^2}{\sigma_z^4}\right) \sum_i^n \left(\mu_{5,i} - 4\,\mu_{3,i}\,\mu_{2,i}\right)$$

Even assuming phenotypic selection described by change in mean and variance alone, models do still not close. For example, change in variance is a function of the skew and kurtosis of the allelic effects. Change in the skew a function of the 4th and 5th order allelic moments, etc.

Genic vs. LD changes

- The previous expressions by Barton & Turelli simply focused on changes in the genic variance (and comparable higher-order moments). Selection-generated LD was ignored
- Conversely, we can ignore changes in the genic variance and examine the effects of LD in changing genetic moments

No assume no allele frequency change, but allow for LD

Barton & Turelli also examined models where only LD contributes to changes, still complex

$$\Delta\mu = \sigma_A^2 L_1 + K_{g,3} L_2 + K_{g,4} L_3 + K_{g,5} L_4$$

$$\Delta\sigma_A^2 = \frac{\sigma_a^2 - \sigma_A^2}{2} - \frac{\Delta\mu^2}{2} + \frac{K_{g,3}}{2} L_1 + \left(\sigma_A^4 + \frac{K_{g,4}}{2}\right) L_2 + \left(3\sigma_A^2 K_{g,3} + \frac{K_{g,5}}{2}\right) L_3 + \left(3K_{g,3}^2 + 4\sigma_A^2 K_{g,4} + \frac{K_{g,6}}{2}\right) L_4$$

$K_{g,x}$ is the jth cumulative of the genotypic distribution, L_i the partial of the mean fitness function with respect to moment i

Where does all this modeling leave us?

- One extreme is the infinitesimal model (no selection-induced changes in allele-frequencies).
- The other extreme is single-locus theory, major changes in allele frequencies
- COA models attempt to bridge these two extreme, and can be thought of as the moderate-term behavior for alleles of modest to small effects.
  - This class of models show how finite population size (N_e) and finite number of loci (n) interact to change the genic variance