

33

Selection in Age-Structured Populations

Draft 26 June 2014

RESPONSE WITH OVERLAPPING GENERATIONS

Many natural populations display age structure, with adults surviving and reproducing over several breeding periods. When generations overlap like this, during any breeding period parents of different ages are likely to have different mean breeding values, and the breeders' equation must be modified. Consider a population under constant directional selection to increase the character. Younger parents are the result of more generations of selection and are thus expected to produce larger offspring than older parents. Thus, while the genotypic value of an individual remains constant over its lifespan, at any given time younger parents have large genotypic values than older parents. This tendency arises from the population structure, and should not be confused with environmental age effects. If such age effects are present, we assume they are either small relative to the genetic change or are predictable enough to be removed before analysis (e.g., Cassuto et al. 1970). We continue to assume that epistasis and genotype \times environment interactions or correlations are negligible.

Asymptotic Response

With overlapping generations, one measure of response to selection is Δ_r , the amount of response seen in newborn offspring per unit time interval (typically years). We use Δ_r to distinguish this *rate of response* from R , the response under a single discrete generation. For comparative purposes, note that the rate of response in unit time intervals for a population with a discrete generation of length τ is $\Delta_r = R/\tau$. The rate of response with overlapping generations was obtained by Dickerson and Hazel (1944) and generalized by Rendel and Robertson (1950), whose derivation we follow. The major assumption is that the amount of genetic gain per unit time interval is constant, leading Hill (1974b) to refer to these as *asymptotic responses*. To reach such a state, a constant amount of selection has to act on an age-structured population for a sufficient amount of time to generate a

stable genetic structure (see Example 4).

We follow the expected change in the population by predicting the change in breeding values. Since parents may contribute differently, the mean parental age for all four possible father-mother/son-daughter combinations, as well as the selection on these combinations, must be considered. Let T_{uv} be the mean parental age for adults of sex u that give rise to sex v . Likewise, let s_{uv} be the genetic superiority of adults of sex u that are used to produce offspring of sex v — the excess in breeding value of these parents relative to unselected adults of the same sex and age. Hence

Parents	Genetic superiority	Mean parental age
Fathers to breed sons	$s_{fa,so}$	$T_{fa,so}$
Fathers to breed daughters	$s_{fa,da}$	$T_{fa,da}$
Mothers to breed sons	$s_{mo,so}$	$T_{mo,so}$
Mothers to breed daughters	$s_{mo,da}$	$T_{mo,da}$

Let g_{so} and g_{da} be the mean breeding value of sons and daughters born in a given year, and Δ_g be the expected change in mean breeding value per year, which is assumed to be constant. Under the assumption of no genotype \times environment interactions, the rate of response $\Delta_r = \Delta_g$. To obtain Δ_g , first consider the breeding value of new-born males. Under the assumption of a constant genetic improvement, the mean breeding values of their fathers is $g_{so} - \Delta_g T_{fa,so} + s_{fa,so}$. This is obtained as follows: the mean breeding value of the population from which fathers were chosen was $g_{so} - \Delta_g T_{fa,so}$, the current mean minus the improvement over the $T_{fa,so}$ years ago that the parents were chosen. The selected fathers are a subsample of this group, with a mean breeding value $s_{fa,so}$ larger than the mean of selected plus unselected fathers. Similarly, their mothers have mean breeding value $g_{da} - \Delta_g T_{mo,so} + s_{mo,so}$. Since parents of both sexes contribute equally, the expected breeding value of sons is the average breeding value of their parents,

$$g_{so} = \frac{1}{2} \left(g_{so} - \Delta_g T_{fa,so} + s_{fa,so} + g_{da} - \Delta_g T_{mo,so} + s_{mo,so} \right) \quad (6.31a)$$

Similarly,

$$g_{da} = \frac{1}{2} \left(g_{so} - \Delta_g T_{fa,da} + s_{fa,da} + g_{da} - \Delta_g T_{mo,da} + s_{mo,da} \right) \quad (6.31b)$$

Equations 6.31a-d have the additional assumption of no fertility differences between parents. Adding these equations together and a little rearrangement gives

$$\Delta_g = \frac{\sum s_{uv}}{\sum T_{uv}} = \frac{s_{fa,so} + s_{fa,da} + s_{mo,so} + s_{mo,da}}{T_{fa,so} + T_{fa,da} + T_{mo,so} + T_{mo,da}} \quad (6.6.)$$

Define the mean generation length (also referred to as the *generation interval*) to be $T = \sum T_{uv}/4$, the average age of a parent. Assuming a linear biparental regression

with all parent-offspring sex combinations having partial regression coefficients $b_{op} = h^2/2$, then $s_{fa,so} = s_{fa,da} = h^2 S_{fa}$ and $s_{mo,so} = s_{mo,da} = h^2 S_{mo}$, giving $\sum s_{uv} = 2h^2(S_{mo} + S_{fa}) = 4h^2 S$. Hence, under these assumptions, the asymptotic rate of response to selection is

$$\Delta_r = \frac{h^2 S}{T} = R/T \tag{6.33}$$

Thus, if the net improvement in the breeding values of each age class remains constant, the mean response per unit time is given by replacing the discrete generation interval τ by the average length with overlapping generations T . There are alternate ways to define the parental age T_{uv} , which require suitable adjustment of the s_{uv} , see James (1977a) for details.

Example 2. Maintaining a roughly constant population size in an experimental population constrains the number of males and females that must be kept as breeding stock, limiting the intensity of selection. Keeping adults for more breeding periods increases both the mean generation length T and the maximal possible selection intensity. Two alternate breeding structures in a cattle population illustrate this point (Hill 1977a). Under population structure I, all offspring born in a given year are sired by two-year old bulls, while their mothers are randomly chosen from females aged 2 to 4 years. Here, $T_{fa,so} = T_{fa,da} = 2$, $T_{mo,so} = T_{mo,da} = (2+3+4)/3 = 3$, giving $T = (2+3)/2 = 2.5$ years. Under population structure II (assuming no mortality), selected bulls are kept an extra year, and females for 2 extra years, giving $T_{fa,so} = T_{fa,da} = (2+3)/2 = 2.5$, $T_{mo,so} = T_{mo,da} = (2+3+4+5+6)/5 = 4$, and $T = (2.5+4)/2 = 3.25$ years. Suppose that under I, at least 7% of the males and all of the females must be used as breeding stock in order to maintain a constant population size. Truncation selection is practiced on a trait with $\sigma_z^2 = 100$ and $h^2 = 0.5$, culling the smallest 93% of the males. The resulting expected selection differential (assuming normality and large population size, using Equation 31.14b) is $\bar{l}_{fa} = 1.9$, giving $s_{fa,so} = s_{fa,da} = h^2 \sigma_z \bar{l}_{fa} = 0.5 \times 10 \times 1.9 = 9.5$. Since there is no selection on mothers, $s_{mo,so} = s_{mo,da} = 0$, and from Equation 6.6. $\Delta_r = (2 \times 9.5)/[(2 \times 3) + (2 \times 2)] = 1.90$. Under population structure II, the fraction of males and females saved at age 2 need only be 50% and 60%, respectively, of structure I to give the same number of breeding adults in any breeding period. Culling all but the largest 3.5% of the two-year old males and culling the smallest 40% of the age two females gives $\bar{l}_{fa} = 2.2$ and $\bar{l}_{mo} = 0.6$. Hence, $s_{fa,so} = s_{fa,da} = 0.5 \times 10 \times 2.2 = 11$ and $s_{mo,so} = s_{mo,da} = 0.5 \times 10 \times 0.6 = 3$, implying $\Delta_r = 2 \times (11 + 3)/[(2 \times 2.5) + (2 \times 4)] = 2.15$, a slightly faster net change than under structure I, even though mean generation time is longer.

Ollivier (1974) has examined the optimal replacement rate of parents to maximize Δ_r given the constraints imposed by population structure. Additional discussions of optimal replacement strategies can be found in James (1977b) and Hopkins and James (1978).

Nonasymptotic Response

Equation 6.6. requires that the rate of genetic change is constant — the mean breeding value of each size and age class increases by Δ_g each year. Clearly, this is often violated. Suppose selection starts in year zero. In year one, only age-one individuals are expected to show genetic change, while by year t , only individuals in age classes one to t have experienced genetic change due to this selection. It can take years for genetic changes to propagate through the population, and many more years before this change is uniformly distributed over the breeding structure (see Example 4). This “lag” in the propagation of genetic change through an age-structured population has economic as well as evolutionary importance. In animal breeding, artificial selection is generally practiced on only a small subset of the entire population. These *nucleus stocks* undergoing selection are always genetically superior to current commercial stocks, which have been bred up from these stocks by several generations of no artificial selection (Richard 1971 discusses this in detail). If the nature of selection changes fairly frequently, the change in mean breeding value in different age-sex classes may never settle down to a constant value and Equation 6.6. is inappropriate.

Searle (1961) was the first to consider how selection-induced changes in breeding values are propagated through an age-structured population. Our treatment follows the more general solution of Hill (1974b, 1977a), which still has the restrictions of no epistasis and no genotype \times environment interactions or correlations. An important limitation of Hill’s approach is that it requires both the covariances between relatives in any two particular age-sex classes and the phenotypic variances in any particular age-sex class to remain constant over time (Pollak 1977). We start by considering how changes in breeding value are propagated over time. Let h and k be the number of years that males and females are kept and $\mathbf{g}(t)$ be the $h + k$ column vector describing mean breeding values at time t . The first h elements correspond to the breeding values of males ages 1 to h , while the $(h + 1)$ - to $(h + k)$ -th elements describe the breeding values of females ages 1 to k . The assumption of constant genetic change required for the asymptotic response Equation 6.6. is that $g_i(t + 1) = g_i(t) + \Delta_g$ for all i and t . More generally, the behavior of $\mathbf{g}(t)$ is obtained using the transition matrix \mathbf{F} ,

$$\mathbf{g}(t) = \mathbf{F}\mathbf{g}(t - 1) = \mathbf{F}^t\mathbf{g}(0) \quad (6.34)$$

Here \mathbf{F}^t denotes the t -th power of \mathbf{F} , not its transpose. \mathbf{F} traces the flow of genes through the age classes, with f_{ij} being the frequency of genes in sex-age class i in time t that come from sex-age class j in time $t - 1$ (e.g., if one out of every four genes in sex-age class 1 come from sex-age class 4, then $f_{14} = 0.25$). Each row of

\mathbf{F} sums to one. In all but the first and $(h + 1)$ -th rows, all elements are zero except for a single element of 1, reflecting the fact that an individual of sex u and age j at time t becomes an individual of sex u and age $j + 1$ at time $t + 1$. These changes due to *aging* are given by the elements

$$f_{i,i-1} = 1 \quad \text{for } i = 2, \dots, h, h + 2, \dots, h + k \quad (6.35a)$$

The first and $(h + 1)$ th rows of \mathbf{F} reflect *reproduction*. The contribution of parents of various ages and sexes to newborn males are given by row 1 and to newborn females by row $h + 1$. Specifically,

$$f_{1,j} = \frac{1}{2} \Pr(\text{male has father of age } j) \quad 1 \leq j \leq h \quad (6.35b)$$

$$f_{1,h+j} = \frac{1}{2} \Pr(\text{male has mother of age } j) \quad 1 \leq j \leq k \quad (6.35c)$$

$$f_{h+1,j} = \frac{1}{2} \Pr(\text{female has father of age } j) \quad 1 \leq j \leq h \quad (6.35d)$$

$$f_{h+1,h+j} = \frac{1}{2} \Pr(\text{female has mother of age } j) \quad 1 \leq j \leq k \quad (6.24e)$$

The expected breeding value of an offspring is the average breeding value of its parents. Hence, the expected breeding value of newborn males at time $t + 1$ is

$$\sum_{j=1}^h f_{1,j} \cdot g_j(t) + \sum_{j=1}^k f_{1,h+j} \cdot g_{h+j}(t)$$

The first sum is the average breeding values of fathers from the various age classes, the second the average breeding value of mothers. Thus, multiplying the first row of \mathbf{F} by $\mathbf{g}_{(t)}$ gives the expected breeding value in male offspring at time $t + 1$. Similarly using the $(h + 1)$ th row of \mathbf{F} gives the expected value for newborn females. The mean age of parents is computed directly from the elements of \mathbf{F} (Hill 1974b), with

$$T_{fa,so} = 2 \sum_{j=1}^h j f_{1,j}; \quad T_{fa,da} = 2 \sum_{j=1}^h j f_{h+1,j} \quad (6.36a)$$

$$T_{mo,so} = 2 \sum_{j=1}^k j f_{1,h+j}; \quad T_{mo,da} = 2 \sum_{j=1}^k j f_{h+1,h+j} \quad (6.36b)$$

While the whole vector \mathbf{g} is monitored, response in overlapping generations is usually measured by the change in the mean character value in *newborn* offspring. With overlapping generations, the vector of breeding values can change simply from the aging of individuals present at the initial generation of selection. These

changes do not directly affect the offspring mean and must be removed. A device for doing this was suggested by Hill (1974b). Distinguish those elements of \mathbf{F} due entirely to aging by defining the matrix \mathbf{Q} , which is \mathbf{F} with all elements in rows 1 and $(h + 1)$ replaced by zero, e.g.,

$$q_{i,i-1} = 1 \quad \text{for } i = 2, \dots, h, h + 2, \dots, h + k \quad (6.37)$$

with all other elements zero. \mathbf{Q}^t gives the distribution of genes initially present at time zero transmitted solely by aging (again, t here denotes the t th power of the matrix, not its transpose). If $t > \max(h, k)$, $\mathbf{Q}^t = \mathbf{0}$ (Hill 1974b), as all individuals present at time zero are dead. The ij th element of \mathbf{F}^t gives the proportion of genes in age-sex class i at year t that are from age-sex class j at time zero, so that $\mathbf{F}^t - \mathbf{Q}^t$ is the proportion of genes due solely to reproduction plus the subsequent aging of these newborns.

To follow how the change in breeding values due to a single generation of selection propagates through the population, assume a single generation of selection at time zero with no selection in subsequent generations. Let the vector \mathbf{s} be the change in breeding values due to selection on each age-sex class. Note that this differs from the selection differential, which measures the change in phenotypic mean (see Example 2, xxx). The expected change in breeding values after one generation is

$$\mathbf{r}(1) = (\mathbf{F} - \mathbf{Q}) \mathbf{s} \quad (6.38a)$$

and after t generations

$$\mathbf{r}(t) = (\mathbf{F}^t - \mathbf{Q}^t) \mathbf{s} \quad (6.38b)$$

Again it should be noted that $\mathbf{r}(t)$ is the resulting response at generation t from selection at generation 0.

Example 3. For population structure I from Example 4, $h = 2$, $k = 3$, and

$$\mathbf{F} = \begin{pmatrix} 0 & 1/2 & 0 & 1/6 & 1/6 & 1/6 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The first row of \mathbf{F} is the proportion of genes in age-one males from the parents in the different age groups. Half of the genes are from fathers, all of whom have age 2. The remaining half are from mothers whose ages are with equal probability two, three, or four. The second row is the two year old males, which obtain all their genes from one year old males, giving $f_{2,1} = 1$. The elements in the remaining

rows follow similarly. Assume that the upper 7% of two-year old males are culled (as in Example 2) at time $t = 0$, in which case (writing the transpose to save space),

$$\mathbf{s}' = (0 \ 9.5 \ 0 \ 0 \ 0 \ 0)$$

Applying Equation 6.38, $\mathbf{r}(t)$, the vector of increments in breeding values at generation t due to a single generation of selection at generation 0, is found to be

$$\mathbf{r}(1) = (\mathbf{F} - \mathbf{Q})\mathbf{s} = \begin{pmatrix} 4.75 \\ 0 \\ 4.75 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{r}(2) = (\mathbf{F}^2 - \mathbf{Q}^2)\mathbf{s} = \begin{pmatrix} 0 \\ 4.75 \\ 0 \\ 4.75 \\ 0 \\ 0 \end{pmatrix}$$

Similarly, $\mathbf{r}(\tau) = (\mathbf{F}^\tau - \mathbf{Q}^\tau)\mathbf{s}$, giving

$$\mathbf{r}(5) = \begin{pmatrix} 2.90 \\ 0.79 \\ 2.90 \\ 0.79 \\ 3.17 \\ 0 \end{pmatrix} \quad \mathbf{r}(10) = \begin{pmatrix} 1.49 \\ 2.39 \\ 1.49 \\ 2.39 \\ 1.6. \\ 2.56 \end{pmatrix} \quad \mathbf{r}(20) = \begin{pmatrix} 1.83 \\ 1.99 \\ 1.83 \\ 1.99 \\ 1.80 \\ 2.02 \end{pmatrix} \quad \mathbf{r}(40) = \begin{pmatrix} 1.90 \\ 1.90 \\ 1.90 \\ 1.90 \\ 1.90 \\ 1.90 \end{pmatrix}$$

In this case, it takes between twenty and forty generations for the changes in breeding values due to a single generation of selection to be uniformly distributed across all age-sex classes. Note that the behavior is rather erratic even after 10 generations. Figure 6.5 plots the change in mean (given by the 1st and 3rd elements in \mathbf{r} , the expected breeding value in newborn males and females) each generation. In the first generation the mean changes by 4.76. This decreases to 0 in the second generation, reflecting the fact that the newborns from selected parents won't have offspring themselves until generation 3. After a sufficient number of generations, the actual change as measured in offspring eventually settles down to the asymptotic value of 1.9 (Example 2). In addition, it takes many generations for the increase in breeding value due to a single generation of selection to be constant across all age groups. Even in generation 10, there is still considerable variation between different age-sex classes and the assumption of a constant genetic gain over all age-sex classes required by Equation 6.31 is clearly violated.

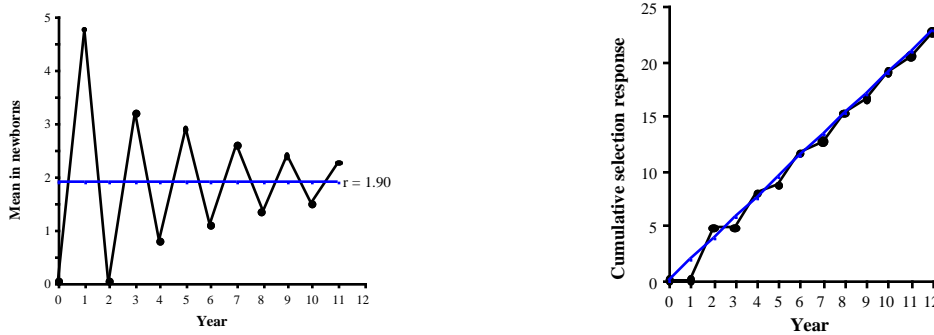


Figure 6.5. The implications of age structure. Age structure I from Example 2 is used, and the population mean before any selection is zero. **Left:** Change in the mean of newborns following a single generation of selection at time $t = 0$. The final change in the mean eventually converges to 1.90, the value predicted from the asymptotic theory. See Example 3 for further details. **Right:** Changes in the mean in newborns following a constant amount of selection applied each generation, starting at $t = 0$. The solid line gives the value predicted from the asymptotic theory. See Example 4 (?) for further details.

Consider now the cumulative response to T generations of selection. Allowing for the possibility of different selection coefficients, let $s(t)$ be changes in breeding value induced by selection at time t . At time T , the response due to selection at $t = 0$ is $(\mathbf{F}^T - \mathbf{Q}^T) \mathbf{s}(0)$, the response due to selection at $t = 1$ is $(\mathbf{F}^{T-1} - \mathbf{Q}^{T-1}) \mathbf{s}(1)$, and in general, the response due to selection at $t = k$ is $(\mathbf{F}^{T-k} - \mathbf{Q}^{T-k}) \mathbf{s}(k)$. Summing these, the cumulative response is

$$\mathbf{R}(T) = \sum_{t=0}^T (\mathbf{F}^{T-t} - \mathbf{Q}^{T-t}) \mathbf{s}(t) \tag{6.39a}$$

If selection is constant,

$$\mathbf{R}(T) = \left(\sum_{t=0}^T (\mathbf{F}^t - \mathbf{Q}^t) \right) \mathbf{s} \tag{6.39b}$$

Example 4. Assume that constant selection is applied each generation with population structure I from Examples 2 and 3. Assume selection occurs by culling all but the largest 7% of one-year old males, giving (again writing the transpose to save space) $\mathbf{s}' = (9.5 \ 0 \ 0 \ 0 \ 0)$. $\mathbf{R}(1) = \mathbf{0}$, as we have to wait one year until

the selected adults can breed. Applying Equation 6.39, the cumulative response following 2, 3, 5, and 25 generations of selection are

$$\mathbf{R}(2) = \begin{pmatrix} 4.75 \\ 0 \\ 4.75 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \mathbf{R}(3) = \begin{pmatrix} 4.75 \\ 4.75 \\ 4.75 \\ 4.75 \\ 0 \\ 0 \end{pmatrix} \quad \mathbf{R}(5) = \begin{pmatrix} 8.71 \\ 7.92 \\ 8.71 \\ 7.92 \\ 4.75 \\ 4.75 \end{pmatrix} \quad \mathbf{R}(25) = \begin{pmatrix} 47.23 \\ 46.37 \\ 47.23 \\ 46.37 \\ 43.42 \\ 41.57 \end{pmatrix}$$

The first and third elements in the vector represent the mean breeding value in newborn males and females, and these are plotted in Figure 6.5 along with $R = 1.90 \times T$, the response predicted from asymptotic theory. Observe that while the per-generation change is often poorly predicted from the asymptotic theory, the cumulative response fits the predictions of asymptotic theory fairly well. While cumulative response may give a reasonable fit, it takes many generations until the stable genetic structure is reached, i.e., where the difference between adjacent age classes are identical. Consider the cumulative response after five generations of selection, $\mathbf{R}(5)$. The difference in mean breeding values between age 1 and age 2 females is $8.71 - 7.92 = 0.79$, between age 2 and age 3 females is $7.92 - 4.75 = 3.17$, and between age 3 and age 4 females is 0. This is a far cry from the constant value of 1.90 predicted from the asymptotic theory. After 25 generations of selection, the corresponding differences are 1.86, 1.95, 1.85, in much better agreement with asymptotic theory.
