

# The Price equation

May 10, 2010

- Classical population genetics is concerned with the change in allele or genotype frequencies.
- Very often, however, we are more interested in the evolution phenotypes.
- Many phenotypes are **quantitative traits**, which can take on a continuous range of values and are influenced by many genes.
- One way to study the evolution of quantitative traits is by multilocus models.
- This requires detailed assumptions about the **genotype-phenotype map**, which is usually unknown empirically.

- An alternative approach is to neglect genetic details and develop models for the evolution of phenotypes directly.
- In the following, we will first study a very general account of evolution by natural selection (the **Price equation**), which can be applied to genotypes as well as phenotypes.
- Then, we will learn about a frame-work called **quantitative genetics**, which is based on simple but often realistic assumptions about the relationship between parents and offspring.

# Variances and covariances

Natural selection depends on genetic variation. Therefore, we will need quantitative measures of this variation. We also need measures of how different traits covary, e.g.

- covariance between different traits of the same individual
- covariance between phenotype and fitness
- covariance between traits in parents and offspring

# The phenotypic variance

Let  $z$  be the value of a phenotypic trait. The variance of  $z$  in a population of  $n$  individuals with trait values  $z_i$  is

$$\begin{aligned}\text{Var}(z) &= \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2 \\ &= \frac{1}{n} \sum_{i=1}^n z_i^2 - \bar{z}^2 = E(z^2) - \bar{z}^2.\end{aligned}$$

If  $p(z)$  is the distribution of  $z$ , then the variance can be computed from

$$\text{Var}(z) = \int_{-\infty}^{\infty} (z - \bar{z})^2 p(z) dz.$$

## Covariance between traits

Let  $x$  and  $y$  be two phenotypic traits. The covariance between  $x$  and  $y$  in a population of  $n$  individuals with traits values  $(x_i, y_i)$  is given by

$$\text{Cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \text{E}(xy) - \bar{x}\bar{y}.$$

Note that  $\text{Cov}(x, x) = \text{Var}(x)$ .

If  $p(x, y)$  is the joint distribution of  $x$  and  $y$ , the covariance can be calculated as

$$\text{Cov}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})(y - \bar{y})p(x, y)dx dy.$$

# Covariance, correlation and regression

The **correlation** between  $x$  and  $y$  is given by

$$\text{Cor}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}}.$$

The correlation is always between  $-1$  and  $1$ .

An important relationship is the following

$$\text{Cov}(x, y) = \beta_{y,x}\text{Var}(x) = \beta_{x,y}\text{Var}(y)$$

Here,  $\beta_{y,x}$  is the slope of the linear regression of  $y$  on  $x$  (i.e. with  $x$  as the independent and  $y$  the dependent variable).

# The Price equation

The Price equation is an extremely general mathematical description of evolutionary change.

- Assume a population of  $n$  individuals with trait values  $Z_1 \dots Z_n$ .
- Each individuals has a certain number  $w_i$  of offspring.
- The offspring phenotypes are given by  $z'_{i,j} = z_i + \delta_{i,j}$ .

The mean trait value in the next generation is given by

$$\begin{aligned}\bar{z}' &= \frac{\sum_{i=1}^n \sum_{j=1}^{w_i} (z_i + \delta_{i,j})}{\sum_{i=1}^n w_i} \\ &= \frac{1}{n\bar{w}} \left( \sum_{i=1}^n w_i z_i + \sum_{i=1}^n w_i \bar{\delta}_i \right) \\ &= \frac{1}{\bar{w}} \left( E(wz) + E(w\bar{\delta}) \right) \\ &= \frac{1}{\bar{w}} \left( \text{Cov}(wz) + \bar{w}\bar{z} + E(w\bar{\delta}) \right)\end{aligned}$$

Hence, the *change* in the mean trait value is

$$\Delta\bar{z} = \frac{1}{\bar{w}} \left( \text{Cov}(w, z) + E(w\bar{\delta}) \right)$$

$$\Delta \bar{z} = \frac{1}{\bar{w}} \left( \text{Cov}(w, z) + E(w\bar{\delta}) \right)$$

- The first term is the covariance between phenotype and fitness. It can arise as a result of selection or genetic drift.
- Assuming drift is absent,  $\text{Cov}(w, z)/\bar{w}$  is the **selection differential**, that is the change in mean phenotype after selection, but before reproduction.
- The second term is the **transmission bias**, that is, the phenotypic change that comes about during reproduction (e.g. due to mutation or recombination), weighted by the parent fitness.

A slight rearrangement of the above equations provides a different perspective:

$$\begin{aligned}\Delta \bar{z} &= \frac{1}{\bar{w}} \left( \text{Cov}(w, z) + E(w\bar{\delta}) \right) \\ &= \frac{1}{\bar{w}} \left( \text{Cov}(w, z) + \text{Cov}(w, \bar{\delta}) + \bar{w}\bar{\bar{\delta}} \right) \\ &= \frac{\text{Cov}(w, z + \bar{\delta})}{\bar{w}} + \bar{\bar{\delta}} \\ &= \frac{\text{Cov}(w, z')}{\bar{w}} + \bar{\bar{\delta}}\end{aligned}$$

Now, the first term is the covariance between parent fitness and offspring phenotype, and the second term ( $\bar{\bar{\delta}}$ ) is the population-wide transmission bias (a systematic change independent of drift and selection).

# The Price equation for allele frequencies

Classical models of allele frequency change can be viewed as special cases of the Price equation.

- First, let's define a generalized “marginal fitness” of a phenotype

$$w_z^* = \frac{1}{N} \sum_{i=1}^N \left( w_i \frac{z_i}{\bar{z}} \right).$$

This is a weighted mean fitness, where individuals with high trait values count more.

- While this may seem not very intuitive, it allows to write

$$\text{Cov}(w, z) = \frac{1}{N} \sum_i w_i z_i - \bar{w} \bar{z} = \bar{z} (w_z^* - \bar{w})$$

- If we define the “trait” of interest as the frequency of an allele in a genotype, then  $w_z^*$  is just the marginal fitness of this allele, and  $\bar{z} = p$  the allele frequency in the population.
- Then, the change in allele frequency is given by

$$\Delta p = \frac{1}{\bar{w}} \left( p(w^* - \bar{w}) + E(w\bar{\delta}) \right).$$

- For example, the basic selection equation for one locus with two alleles can be written as

$$\Delta p = \frac{p}{\bar{w}} (w^* - \bar{w}).$$

- In a two-locus model with two alleles per locus, we can focus on haplotype frequencies.
- Earlier we have seen that haplotype frequencies change according to

$$\Delta x_{ij} = \frac{1}{\bar{w}} \left( x_{ij} (w_{ij}^* - \bar{w}) \pm r D w_{1122} \right),$$

which has the same structure as the Price equation.

- Here the transmission bias term represents the change in haplotype frequencies due to recombination.

# The Price equation and multilevel selection

Selection at more than one level of biological organisation has been a very controversial topic for decades. The Price equation offers one way to conceptualize it.

- Assume a population composed of several subpopulations.
- Define the phenotype and fitness of a subpopulation,  $Z$  and  $W$ , as the respective means over the individuals in the subpopulation.
- This is an example of “multilevel-selection 1”, which can also be seen, more traditionally, as frequency-dependent selection in a structured population.

- Ignoring the population structure, the change in mean phenotype in the total population is

$$\Delta \bar{z} = \frac{1}{\bar{w}} \left( \text{Cov}(w, z) + E(w\bar{\delta}) \right).$$

- The covariance term can be separated into a within-subpopulation and a between-subpopulation component (as in analysis of variance):

$$\text{Cov}(w, z) = \text{Cov}(W, Z) + E_k(\text{Cov}(w, z))$$

where  $E_k$  stands for the expectation over subpopulations.

- These two components can be interpreted as describing selection on the individual and subpopulation-level, respectively.

- Alternatively, we can focus on the evolution of the mean subpopulation phenotype  $\bar{Z}$

$$\Delta\bar{Z} = \frac{1}{\bar{W}} \left( \text{Cov}(W, Z) + E(W\bar{\delta}_g) \right).$$

- Here, selection on individuals within subpopulations can be interpreted as part of the transmission bias  $\bar{\delta}_g$  at the subpopulation level.

# Fisher's fundamental theorem

By treating fitness itself as the trait of interest, the second form of the Price equation becomes

$$\Delta \bar{w} = \frac{\text{Cov}(w, w')}{\bar{w}} + \bar{\delta}(w)$$

This is a form of **Fisher's fundamental theorem of natural selection**:

- The first term is the change in mean fitness due to selection. Parent-offspring covariance should be positive, so selection always increases mean fitness.
- The second term is what Fisher called the **deterioration of the environment**. It is an overall decrease in mean fitness due to, e.g., frequency-dependent selection or coevolution.