

6 Selection in quantitative characters

There are several levels of population description. At the most fundamental level, we describe all genotypes represented in the population. With two alleles at each of L loci, one needs 2^{2L} dynamic variables to describe the population. In some situations (for example if the population is at Hardy-Weinberg proportions), one can use a simpler description based on 2^L gamete frequencies instead of genotype frequencies. Invoking further assumptions (e.g. linkage equilibrium assumption justified by assuming that selection is weak) one can simplify the approach even more by using allele frequencies. This gives only L dynamic variables. Next, we consider the simplest approach for describing populations based on using a single dynamic variable representing the average value of a quantitative character.

6.1 Standard model for a quantitative character

Quantitative traits are phenotypic traits that exhibit continuous variation and are subject to microenvironmental effects. Examples: size, weight etc. Quantitative traits are thought to be controlled by *many* loci with *small* effects.

The standard model for a quantitative character is described by equation

$$z = g + e,$$

where z is the trait value, g is the contribution of genotype (“genotypic value”; the average trait value for a group of organisms with the same genotype), and e is the contribution of microenvironment. First, we will assume that g and e are independent (that is there is no genotype-environment interaction). *Genotypic value* g can be thought of as a sum of contributions from many loci:

$$g = \sum(\alpha_i + \alpha'_i),$$

where α_i and α'_i are the contributions of the i -th locus from paternal and maternal gamete, respectively. The above model implies that the trait is additive. *Microenvironmental deviation* e is usually modeled as a random variable with zero mean and a constant variance:

$$\begin{aligned}\bar{e} &= 0, \\ \text{var}\{e\} &= E.\end{aligned}$$

Population state is described by the *genotypic distribution*, $p(g)$, of g in the population. One can also define the moments: the mean value, \bar{g} , variance, G , etc.:

$$\begin{aligned}\bar{g} &= \int gp(g)dg, \\ G &= \int (g - \bar{g})^2 p(g)dg.\end{aligned}$$

Phenotypic distribution, $p(z)$, has the mean, \bar{z} , and the phenotypic variance, P :

$$\begin{aligned}\bar{z} &= \bar{g}, \\ P &\equiv \text{var}\{z\} = G + E.\end{aligned}$$

6.2 Types of viability selection

In major locus models that we studied before fitness was usually assigned to genotype, e.g. w_{AA} is fitness of genotype **AA** (genotype \Rightarrow fitness). With quantitative traits fitness depends on phenotype which in turn is controlled by genotype and environment (genotype + environment \Rightarrow phenotype \Rightarrow fitness).

Examples of phenotypic fitness function, $w = w(z)$.

Directional selection:

$$w = a + bz \text{ or } w = e^{az}.$$

Here, fitness function monotonically increases (or decreases) with z .

Stabilizing selection:

$$w = 1 - sx^2 \text{ or } w = e^{-az^2}.$$

Here, fitness reaches a maximum at an intermediate value of z .

Disruptive selection:

$$w = e^{-a(z-1)^2} + e^{-a(z+1)^2}.$$

Here, fitness increases with deviation from an intermediate value of z .

The mean fitness of the population is defined as

$$\bar{w} = \int w(z)p(z)dz = \int w(g)p(g)dg,$$

where the (induced) genotypic fitness is

$$w(g) = \int w(g+e)p(e)de.$$

6.3 Robertson-Price formula

If $p(z)$ is the phenotypic distribution before selection, then the phenotypic distribution after selection is

$$p_s(z) = \frac{w(z)}{\bar{w}}p(z).$$

Let $\psi = \psi(z)$ be a function of z . The mean value of ψ before selection is

$$\bar{\psi} = \int \psi(z)p(z)dz.$$

After selection

$$\bar{\psi}_s = \int \psi(z)p_s(z)dz = \int \psi(z)\frac{p(z)w(z)}{\bar{w}}dz.$$

The change in $\bar{\psi}$ as a result of selection is

$$\Delta\bar{\psi} \equiv \bar{\psi}_s - \bar{\psi} = \int \psi(z)\frac{p(z)w(z)}{\bar{w}}dz - \bar{\psi} = \frac{\int[\psi(z)w(z)]p(z)dz - \bar{\psi}\bar{w}}{\bar{w}} = \frac{\overline{\psi w} - \bar{\psi}\bar{w}}{\bar{w}}.$$

Thus,

$$\Delta\bar{\psi} = \frac{cov(\psi, w)}{\bar{w}}, \tag{25}$$

where $cov(a, b)$ is the covariance of a and b . This is the Robertson-Price formula. Note that no specific assumptions about the distributions and selection regimes have been made so far.

For example, if $\psi = z$, then

$$\Delta\bar{z} = \frac{cov(z, w)}{\bar{w}}. \tag{26}$$

If $\psi = z^2$, then

$$\Delta\bar{z}^2 = \frac{cov(z^2, w)}{\bar{w}}. \tag{27}$$

The Robertson-Price formula predicts the change in a specific population characteristic as a result of within-generation selection. Under some additional assumptions it can be used to describe the changes between generations. For example, if the trait z is additive, then recombination and segregation will not change its mean value. Thus, equation (26) describes the change in the mean trait value between two subsequent generations.

Homework: Assume that the distribution of z before selection has mean \bar{z} , variance P and third moment μ_3 ($\mu_3 \equiv \int (z - \bar{z})^3 p(z) dz$). Find $\Delta\bar{z}$ for linear and quadratic fitness functions ($w_{lin} = a + bz, w_{quad} = 1 - sz^2$).

6.4 Lande formula: normal approximation

Let us assume that both the distribution of g and the distribution of e are Gaussian:

$$p(g) = \frac{1}{\sqrt{2\pi G}} \exp\left(-\frac{(g - \bar{g})^2}{2G}\right),$$

$$p(e) = \frac{1}{\sqrt{2\pi E}} \exp\left(-\frac{e^2}{2E}\right).$$

The distribution of z will be normal as well:

$$p(z) = \frac{1}{\sqrt{2\pi P}} \exp\left(-\frac{(z - \bar{z})^2}{2P}\right),$$

where $P = G + E, \bar{z} = \bar{g}$. Differentiating \bar{w} with respect to \bar{g}

$$\begin{aligned} \frac{\partial \bar{w}}{\partial \bar{g}} &= \int w(g) \frac{\partial p(g)}{\partial \bar{g}} dg = \int w(g) \frac{g - \bar{g}}{G} p(g) dg \\ &= \frac{1}{G} \left[\int w(g) g p(g) dg - \bar{g} \int w(g) p(g) dg \right]. \end{aligned}$$

Thus,

$$\frac{1}{\bar{w}} \frac{\partial \bar{w}}{\partial \bar{g}} = \frac{1}{G} \left[\int g \frac{w(g)p(g)}{\bar{w}} dg - \bar{g} \right] = \frac{1}{G} (\bar{g}_s - \bar{g}) = \frac{1}{G} R, \quad (28)$$

where $R \equiv \bar{g}_s - \bar{g}$ is *selection response*.

In a similar way, differentiating \bar{w} with respect to \bar{z}

$$\begin{aligned} \frac{\partial \bar{w}}{\partial \bar{z}} &= \int w(z) \frac{\partial p(z)}{\partial \bar{z}} dz = \int w(z) \frac{z - \bar{z}}{P} p(z) dz \\ &= \frac{1}{P} \left[\int w(z) z p(z) dz - \bar{z} \int w(z) p(z) dz \right]. \end{aligned}$$

Thus,

$$\frac{1}{\bar{w}} \frac{\partial \bar{w}}{\partial \bar{z}} = \frac{1}{P} \left[\int z \frac{w(z)p(z)}{\bar{w}} dz - \bar{z} \right] = \frac{1}{P} (\bar{z}_s - \bar{z}) = \frac{1}{P} S, \quad (29)$$

where $S \equiv \bar{z}_s - \bar{z}$ is *selection differential*.

Combining (28) and (29),

$$R = \frac{G}{P} S = h^2 S, \quad (30)$$

where

$$h^2 = \frac{G}{G + E}$$

is *heritability* (in the broad sense). Heritability characterizes the proportion of heritable genetic variation in the overall phenotypic variation. Equation (30) is known as the breeders' equation. It shows that selection response equals heritability times selection differential.

From (28),

$$R = G \frac{\partial \ln \bar{w}}{\partial \bar{g}} \quad (31)$$

For an additive trait, segregation and recombination do not change the mean trait value ($\bar{g}' = \bar{g}_s$, $R = \Delta \bar{g} \equiv \bar{g}' - \bar{g}$). Thus, the change in \bar{g} *between generations*

$$\Delta \bar{g} = G \frac{\partial \ln \bar{w}}{\partial \bar{g}} \quad (32)$$

(Lande, 1976). If genotypic variance G does not change, one can use (32) for predicting the long-term dynamics. Because $\bar{z} = \bar{g}$ and $\partial \ln \bar{w} / \partial \bar{g} = \partial \ln \bar{w} / \partial \bar{z}$, one can also write

$$\Delta \bar{z} = G \frac{\partial \ln \bar{w}}{\partial \bar{z}}. \quad (33)$$

Implications: gradient-type dynamics (evolution towards an equilibrium, no cycles, no chaos, average fitness is maximized).

6.5 Lande formula: multivariate case

Let there be n phenotypic traits:

$$z_i = g_i + e_i, \quad i = 1, \dots, n.$$

Assume that the distribution of the genotypic values g_i is multivariate normal with the mean $\bar{g} = (\bar{g}_1, \dots, \bar{g}_n)^T$ and a $n \times n$ covariance vector G . Then the change in \bar{g} in one generation is

$$\Delta \bar{g} = G \frac{\partial \ln \bar{w}}{\partial \bar{g}},$$

where the vector of selection gradients $\frac{\partial \ln \bar{w}}{\partial \bar{g}} = (\frac{\partial \ln \bar{w}}{\partial \bar{g}_1}, \dots, \frac{\partial \ln \bar{w}}{\partial \bar{g}_n})^T$ (Lande, 1979).

For example in the case of two traits x and y

$$\begin{pmatrix} \Delta \bar{x} \\ \Delta \bar{y} \end{pmatrix} = \begin{pmatrix} G_x & C \\ C & G_y \end{pmatrix} \begin{pmatrix} \frac{\partial \ln \bar{w}}{\partial \bar{x}} \\ \frac{\partial \ln \bar{w}}{\partial \bar{y}} \end{pmatrix}, \quad (34)$$

where G_x and G_y are genotypic variances, and C is covariance of x and y . Note that even if a trait, say y , does not affect fitness ($\partial \ln \bar{w} / \partial \bar{y} = 0$), it will change if $C \neq 0$ (correlated response to selection).

6.6 Lande formula: weak selection approximation

The assumption that all relevant distributions stay normal is not easy to justify. Also, the mean fitness of the population can be found easily only in some special cases. Here, we consider an alternative method of the derivation of equations analogous to (32) based on less restrictive assumptions.

We start with the Robertson-Price formula

$$\Delta \bar{g} = \frac{\text{cov}(g, w)}{\bar{w}}. \quad (35)$$

Expanding $w = w(g)$ in a Taylor series at $g = \bar{g}$, one gets

$$w(g) = w(\bar{g}) + \frac{dw}{dg}(g - \bar{g}) + \frac{1}{2} \frac{d^2w}{dg^2}(g - \bar{g})^2 + \dots \quad (36)$$

Using the covariance properties, we find that

$$\begin{aligned} \text{cov}(g, w(g)) &= \text{cov}(g - \bar{g}, w(g)) \\ &= \text{cov}(g - \bar{g}, w(\bar{g}) + \frac{dw}{dg}(g - \bar{g}) + \frac{1}{2} \frac{d^2w}{dg^2}(g - \bar{g})^2 + \dots) \\ &= 0 + \frac{dw}{dg} \text{cov}(g - \bar{g}, g - \bar{g}) + \frac{1}{2} \frac{d^2w}{dg^2} \text{cov}(g - \bar{g}, (g - \bar{g})^2) + \dots \\ &= G \frac{dw}{dg} + \frac{1}{2} \frac{d^2w}{dg^2} \mu_3 + \dots, \end{aligned}$$

where $\mu_3 = \int (g - \bar{g})^3 p(g) dg$ is the third moment of the distribution of g (which measures asymmetry of $p(g)$) and all derivatives are evaluated at $g = \bar{g}$.

Computing the expectation of both sides of (36), one can see that

$$\begin{aligned} \bar{w} &= w(\bar{g}) + \frac{dw}{dg}(g - \bar{g}) + \frac{1}{2} \frac{d^2w}{dg^2}(g - \bar{g})^2 + \dots \\ &= w(\bar{g}) + \frac{1}{2} G \frac{d^2w}{dg^2} + \dots \end{aligned}$$

Thus,

$$\Delta \bar{g} = \frac{G \frac{dw}{dg} + \frac{1}{2} \frac{d^2w}{dg^2} \mu_3 + \dots}{w(\bar{g}) + \frac{1}{2} G \frac{d^2w}{dg^2} + \dots}$$

which can be approximated by

$$\Delta \bar{g} = G \frac{d \ln w}{dg}, \quad (37)$$

where the derivative is evaluated at $g = \bar{g}$. Note that to apply (37) one needs to know the derivative of the individual fitness rather than the mean fitness of the population. The approximation is good is

$$G \frac{d^2w}{dg^2} \ll w(\bar{g}), \quad \mu_3 \frac{d^2w}{dg^2} \ll G \frac{dw}{dg}.$$

The former condition is satisfied if differences between fitnesses are small (weak selection). The latter condition is satisfied if differences between selection gradients, dw/dg , are small (weak non-linearity in selection; if $w(g)$ is linear, all derivatives higher than first will be zero). Note that equation (37) can be used even if individual fitness depends on the state of the population (that is with frequency dependent fitness, e.g. $w = w(z, \bar{z})$).

6.7 Mutation-selection balance

Mutations are random changes in the genetic material of a cell. Most mutations are harmful, but some are advantageous. Mutations occur naturally at low rates ($10^{-5} - 10^{-6}$ per gene per generation). Mutations are the ultimate source of genetic variation.

We will consider a populations of organisms different with respect to a single locus with two alleles, **A** and **a**. We assume that fitnesses are $w_A = 1$ (allele **A** is normal) and $w_a = 1 - s$ (allele **a** is deleterious, $s > 0$). Let p and q be the frequencies of the alleles ($p + q = 1$).

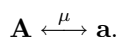
We already know that *after selection*:

$$p' = \frac{w_A}{\bar{w}} p, \quad q' = \frac{w_a}{\bar{w}} q,$$

and that the change in the frequency of allele **A** is

$$\Delta_s p = \frac{pq(w_A - w_a)}{\bar{w}} = \frac{spq}{\bar{w}} (\geq 0).$$

We assume that alleles can mutate with a small probability μ ($\ll 1$) which is the same for forward and backward mutations:



The allele frequency *after mutation* is

$$p'' = (1 - \mu)p' + \mu q' = p' + \mu(q' - p').$$

Thus the change per generation is

$$\Delta p \equiv p'' - p = p' - p + \mu(q' - p') = \frac{spq}{\bar{w}} + \mu \frac{w_a q - w_A p}{\bar{w}} = \frac{spq + \mu(-sq + q - p)}{\bar{w}}. \quad (38)$$

At equilibrium (that is at a balance of mutation and selection) $\Delta p = 0$ and

$$spq - \mu sq + \mu(q - p) = 0$$

which can be solved exactly. (This is a quadratic equation in p .) Instead of the exact solution, however, one can use a simple approximation. Because μ is small (say, $\approx 10^{-6}$), we expect q to be small as well, $q - p = 2q - 1 \approx -1$, $pq = (1 - q)q \approx q$. Thus, we expect $sq \approx \mu$ and

$$q^* \approx \frac{\mu}{s}, \quad (39)$$

which is the frequency of the deleterious allele **a** maintained in the population by mutation in spite of selection. The frequency of the deleterious allele will be rather small though. For example, with $s = .01$ and $\mu = 10^{-6}$, $q^* = 10^{-5}$.

Using a weak selection approximation (that is assuming that s is small so that $\bar{w} \approx 1$), equation (38) can be approximated by a (second order ordinary) differential equation

$$\dot{p} = -spq + \mu(q - p),$$

which can be solved exactly.

6.7.1 The Mutation load

The change of average fitness associated with maintaining the variability in a population was called the *genetic load* by Muller (1950). To Muller the load was a burden, measured in terms of the reduced fitness, but felt in terms of death, sterility etc. In most evolutionary considerations, however, it is used as a measure of the amount of natural selection associated with a certain amount of genetic variation. The genetic load is usually defined as

$$L = \frac{w_{max} - \bar{w}}{w_{max}},$$

where w_{max} is the maximum possible fitness.

Using our results above, at the mutation-selection balance equilibrium

$$w_{max} = 1, \bar{w} = (1 - q^*) \times 1 + q^* \times (1 - s) = 1 - \mu$$

so that the mutation load (that is the genetic load resulting from mutations maintaining deleterious alleles) is

$$L = \mu.$$

Surprisingly, this load does not depend on s . Thus, increasing the rate of even slightly deleterious mutations (with very small s) can be very harmful!

6.7.2 Major points

- mutation introduces and maintains genetic variation;
- the mutation load due to the deleterious mutations does not depend of the strength of selection against them.

7 Random genetic drift

So far in describing evolutionary dynamics we have been using a *deterministic* framework allowing for no stochastic fluctuations of any kind. Using this framework, however, requires some additional assumptions. In particular, our description of different populations in terms of allele frequencies implies that the population size is very large. In this and the next two sections, we will consider how allowing for stochastic effects resulting from the finiteness of biological populations affects evolutionary dynamics.

An important notion in evolutionary population genetics is that of *random genetic drift* by which one usually means *random* variation in genotype frequencies resulting from the finiteness of populations. There are two important effects of random genetic drift that we will consider. Genetic drift (i) removes genetic variation from populations and (ii) affects the probability of survival of new mutations.

We consider a *finite population* of size N assuming discrete and non-overlapping generations. There is a single locus with k alleles which are identical with respect to viability and fertility (i.e. there is no selection). We assume that each adult produces a very large number of offspring of which only N (randomly chosen from a common offspring pool) organisms will develop into the adults of the next generation. Let us consider an allele \mathbf{A} that has frequency p (so that on average there are Np alleles \mathbf{A} out of N alleles present in the population). Our assumptions about the way the next generation is formed translates into assuming that the probability that i alleles \mathbf{A} made it to the next generation is given by the binomial probability

$$\text{Prob}\{\text{there are } i \text{ alleles } \mathbf{A}\} = \frac{N!}{i!(N-i)!} p^i (1-p)^{N-i},$$

where i can be $0, 1, \dots, N$. This is a so called *Fisher-Wright binomial scheme*.

Of course, changes in allele frequencies in any individual population resulting from random genetic drift cannot be predicted. But the *average behavior* in a large number of populations can be predicted. Let p be the frequency of allele \mathbf{A} in this generation. We consider the expected frequency, $E\{p'\}$, of this allele in the next generation. The mean of a binomial variable with parameters N and p is Np so that

$$E\{p'\} \equiv E\{i/N\} = p.$$

This shows that the expected change in allele frequency is zero, $E\{\Delta p\} = 0$. The variance of a binomial variable with parameters N and p is $Np(1-p)$ so that the variance in p' ,

$$\text{var}\{p'\} \equiv \text{var}\{i/N\} = \frac{pq}{N},$$

where $q = 1 - p$. This shows that the variance of the allele frequency change is $\text{var}\{\Delta p\} = \frac{pq}{N}$ and is positive. Thus, the allele frequencies are expected to change.

[Random genetic drift: fluctuations in the genetic structure of a population (which is characterized by allele/genotype frequencies) resulting from inherent stochasticity in the process of reproduction (number of offspring, segregation, recombination, mutation etc) and the finiteness of the population size.

One locus multiallele haploid system: if each individual leaves the same number of offspring, no drift even if the population size is finite.]

We consider a one-locus haploid population. Let f_t be the probability that two randomly chosen distinct individuals in generation t are *identical in state* (that is carry alleles of the same type). Note that $h_t = 1 - f_t$ can be interpreted as a measure of genetic variability of the population. (With only two alleles with frequencies p and q , $f_t = p_t^2 + q_t^2, h_t = 2p_tq_t$).

Figure ? will be here

Let us consider the next generation. Two randomly chosen distinct individuals are identical in state if (i) they have the same “parent” (which happens with probability $1/N$) or (ii) if they have different parents (which happens with probability $1 - 1/N$) but these parents already were identical in state (which happens with probability f_t). Thus, the dynamics of f_t can be described by a single recurrence equations

$$f_{t+1} = \frac{1}{N} + (1 - \frac{1}{N})f_t.$$

The corresponding equation for h_t is even simpler:

$$h_{t+1} = 1 - f_{t+1} = (1 - \frac{1}{N})h_t.$$

The solution of this equation is

$$h_t = h_0(1 - \frac{1}{N})^t \approx h_0e^{-t/N}$$

where the approximation is good if N is sufficiently large. Thus, as $t \rightarrow \infty$, $h_t \rightarrow 0$ and only one allele remains in the population. The characteristic time for approaching a monomorphic state is approximately N generations.

Note that $h \rightarrow 0$ independently of how many alleles are present initially. Even if initially all N individuals have different alleles ($k = N$), in N generations $h \approx 0$. Another way to put it is to say that all alleles present right now can be traced to a single ancestor allele $\approx N$ generations ago! This also means that with no differences in fitness, the probability of fixation of a specific allele is $1/N$.

7.0.3 Major points

- random genetic drift removes genetic variability;
- the characteristic time of fixation is order N generations.

Homework:

What is the probability that a particular allele gets at least a copy in the next generation? Hint: consider the probability of the complementary event (=“no copies of this allele in the next generation”).

7.1 Drift and mutation

With mutation the permanent loss of genetic variation is, obviously, impossible. Here we consider how including mutation affects the dynamics of genetic variation. We assume that there are k alleles \mathbf{A}_i ($i = 1, \dots, k$) and that each allele can mutate (after reproduction) to any other $k - 1$ alleles with probability $\mu/(k - 1)$:

$$\mathbf{A}_i \xleftrightarrow{\frac{\mu}{k-1}} \mathbf{A}_j.$$

Note that the overall probability that an allele mutates is μ .

Figure ? will be here

The probability of identity in state after reproduction but before mutation satisfies to an expression derived in the previous section:

$$f'_{t+1} = \frac{1}{N} + (1 - \frac{1}{N})f_t. \quad (40a)$$

The probability of identity in state in the next generation (after reproduction and mutation) is

$$f_{t+1} = f'_{t+1} [(1 - \mu)^2 + \mu^2] + (1 - f'_{t+1}) 2 \frac{\mu}{k-1} (1 - \mu). \quad (40b)$$

Here the first term in the right-hand side specifies the probability that two randomly chosen allele were identical before mutation and did not mutate or did mutate but to the same allele. The second term specifies the probability that the allele that were different in state after reproduction become identical after mutation in one of them.

Equations (40) can be used to show that as time increases, f_t approaches an equilibrium value which can be approximated (assuming that $\mu \ll 1$) as

$$f = \frac{2N\mu \frac{1}{k-1} + 1}{2N\mu \frac{k}{k-1} + 1}. \quad (41a)$$

This simplifies to

$$f = \frac{2N\mu + 1}{4N\mu + 1} \quad (41b)$$

for $k = 2$, and to

$$f = \frac{1}{2N\mu + 1} \quad (41c)$$

for $k \rightarrow \infty$. Note that $N\mu$ is the number of mutants per generation. With $N\mu = 1$, $f = 3/5$ for $k = 2$ and $f = 1/3$ for large k . Thus, a small number of mutants per generation is sufficient to maintain large levels of genetic variability in a finite population (with no selection). If $N\mu \ll 1$ (which is the case if the population size is very small), then $f \approx 1$ (that is the level of variability is very small).

7.1.1 Major points

- mutation maintains genetic variation in finite populations;
- an important parameter controlling the level of genetic variation maintained is the number of mutants per generation, $N\mu$.

7.2 Drift and selection

We already know that the probability of fixation of a neutral allele in a finite population of size N is $1/N$. Here, we derive the probability of fixation of a selected allele.

7.2.1 Derivation of the fixation probability (Diffusion approximation)

Let $u(p)$ be the ultimate probability of fixation of allele \mathbf{A} given that its initial frequency is p . Then, using Markovian nature of the process of change in p , one can write

$$u(p) = \sum_{\Delta p} \text{Prob}(\Delta p) u(p + \Delta p), \quad (42)$$

where $\text{Prob}(\Delta p)$ is the probability of a particular change in p in one generation, and $u(p + \Delta p)$ is the probability of fixation given that this particular change has happened. Equation (42) can also be thought of as a variant of the generalized partition theorem in the probability theory: $\text{Prob}(A) = \sum_i \text{Prob}(A|B_i)\text{Prob}(B_i)$, where $\text{Prob}(A)$ is the probability of event A and $\text{Prob}(A|B_i)$ is the conditional probability of A given B_i . The expression in the right-hand side of (42) can also be interpreted as the mathematical expectation of $u(p + \Delta p)$ with respect to all possible changes Δp , allowing one to write

$$u(p) = E_{\Delta p}(u(p + \Delta p)).$$

Expanding $u(p + \Delta p)$ in a Taylor series

$$u(p + \Delta p) \approx u(p) + u'(p)\Delta p + \frac{1}{2}u''(p)(\Delta p)^2,$$

one finds that

$$u(p) \approx u(p) + u'(p)E[\Delta p] + \frac{1}{2}u''(p)E[(\Delta p)^2].$$

We assume that $E\{(\Delta p)^2\} = \text{var}\{\Delta p\} + (E\{\Delta p\})^2 \approx \text{var}\{\Delta p\}$ that is that the expected change in allele frequency is small. Let $m(p)$ and $v(p)$ be the expected change in p and the expected variance of the change in p . Then, $u(p)$ satisfies to a linear homogeneous second order ODE

$$m(p)u'(p) + \frac{1}{2}v(p)u''(p) = 0 \quad (43)$$

with boundary conditions $u(0) = 0, u(1) = 1$. The solution can be found explicitly:

$$u(p) = \frac{\int_0^p G(x)dx}{\int_0^1 G(x)dx}, \quad (44)$$

where

$$G(x) = e^{-\int \frac{2m(x)}{v(x)} dx}.$$

To find this solution we can use *Maple*. We also can do in an old-fashioned way: first, convert equation (43) to a first order ODE $f' + (2m/v)f = 0$ by changing the dependent variable to $f(p) = u'(p)$, second, compute the integrating factor $H = \exp(\int \frac{2m}{v} dp)$, third, solve the ordinary differential equation $d(fH)/dp = 0$ for f , and, finally, integrate the answer to find p subject to the boundary conditions.

7.2.2 Example: the probability of fixation of an advantageous allele

We consider a one-locus two-allele haploid populations of finite size N . Let the fitnesses of alleles **a** and **A** be 1 and $1 + s$, respectively. The mean and variance of the change in the frequency p of allele **A** in one generation are

$$m(p) = spq, \quad v(p) = pq/N,$$

where $q = 1 - p$. Substituting these values into the general expression (44) one finds that the probability of fixation is

$$u(p) = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}}.$$

Assuming that there is a single copy of the mutant allele initially ($p = 1/N$),

$$u(1/N) = \frac{1 - e^{-2s}}{1 - e^{-2Ns}}. \quad (45)$$

If s is small (weak selection) but Ns is large (large population),

$$u \approx 2s,$$

that is the fixation probability is approximately twice the selective advantage. In large populations, selection will overwhelm drift once the advantageous allele is at all common. But only a very small proportion of advantageous mutations have a chance to become common!

7.2.3 Example: the probability of fixation of a deleterious allele

For a deleterious allele ($s < 0$)

$$u \approx \frac{e^{2|s|} - 1}{e^{2N|s|} - 1},$$

which can be large enough if Ns is small. Thus, in small populations, the fixation of deleterious alleles can occur with a non-negligible probability! If Ns is large, the probability of fixation $u \approx 2s/e^{2Ns}$ and is very small.

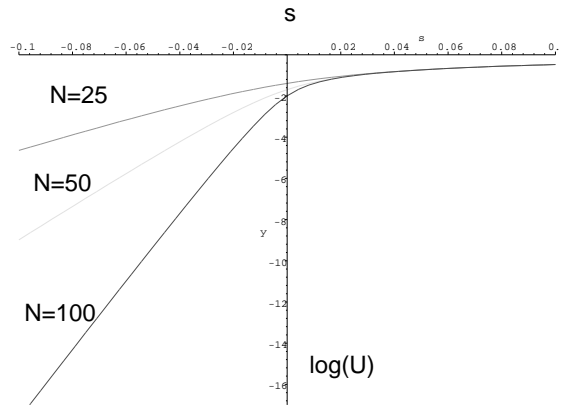


Figure 3: Fixation probability (45) as function of the allele effect s and the population size N .

Homework. Consider a one-locus two-allele model with frequency-dependent fitnesses

$$w_A = 1 + sp,$$
$$w_a = 1 + sq,$$

describing selection against rare alleles. In the deterministic approximation, the corresponding dynamic system has two locally stable monomorphic equilibria (corresponding to fixation of alleles). Finite populations subject to mutation can “jump” between these equilibria. These “jumps” will happen if an allele that has a low frequency initially gets fixed. Try to show that the probability of fixation of allele **A** that is represented by a single copy in a population of size N can be approximated (if $N \gg 1$ and $Ns > 4$) as

$$U = e^{-Ns/2} \sqrt{2s/N} \frac{1}{\sqrt{\pi}}.$$

Numerically compare U with the probability of fixation of a neutral allele for a set of parameter values (for example, with $N = 10, 100, 1000$ and $s = 0.1, 0.01, 0.001$).

7.2.4 Major points

In a finite population

- the probability of fixation of an advantageous allele is approximately twice its selective advantage;
- the fixation of slightly deleterious alleles can occur with a non-negligible probability;
- “jumps” between the corresponding locally stable deterministic equilibria are possible.