MSc Computational Methods in Ecology and Evolution

Evolutionary Modelling

Dr Bhavin S. Khatri (bkhatri@imperial.ac.uk)

Tutorial – 18th Feb 2021;

Question 1   Time to fixation for neutral mutants

In the lectures we derived that the time to fixation for pure random genetic drift (no selection, no mutation), was proportional to the population size $\tau_{fix} \sim N$.

a) A calculation for the time to fixation, or loss, using the diffusion approximation that accounts also for the initial frequency x_0 gives
$$\tau = -2N(x_0 \ln x_0 + (1 - x_0) \ln(1 - x_0))$$
i) plot $\tau/2N$, ii) what value of x_0^* maximises the time? iii) explain intuitively (e.g. by symmetry) why there is a maximum at this value x_0^*.

b) Plot τ now as a function of N for $x_0 = 0.01, 0.1, 0.5$ and $100 \leq N \leq 10^8$. Use this plot to argue that changes in population size have a much greater effect on τ than changing the initial frequency.

c) If the initial frequency arises as a de novo mutation $x_0 = 1/N$. Show, for $N \gg 1$, that in this case
$$\tau \approx 2 \ln N.$$
What has happened to the linear dependence on N above?

d) If a mutant is under selection, we showed in the lectures that the establishment time was approximately, $\tau_{est} \approx \frac{1}{2s}$. At establishment, what will be the frequency of the mutant?

e) Using this result, show a rough estimate of the time to fixation, once the mutant has established is
$$\tau_{fix} \approx \frac{1}{s} \ln(2Ns)$$
assuming the deterministic approximation $t^* = \frac{1}{s} \ln\left(\frac{1}{x_0} - 1\right)$, where t^* is the time for a mutant to reach frequency $\frac{1}{2}$, assuming a deterministic model of selection. Is this an underestimate or overestimate of the time to fixation?

f) Putting this together with the time to establishment show that including establishment the time to fixation of a new mutant is approximately
$$\tau_{fix} \approx \frac{1}{s}\left(\frac{1}{2} + \ln(2Ns)\right)$$

g) i) Which factor, selection or drift is dominant in determining the time to fixation and explain your reasoning? ii) What else is this calculation ignoring?

Question 2   Malthusian selection equations

A continuous time description of selection begins with defining *Malthusian* growth rates for wild type and mutant v_1 and v_2, respectively. Each population grows as

$$\frac{dn_1}{dt} = v_1 n_1; \quad \frac{dn_2}{dt} = v_2 n_2.$$

a) If the frequency of the mutant is $x = n_2/(n_1 + n_2)$, using this change of variable and the above ODEs, show that the resulting differential equation for x is

$$\frac{dx}{dt} = (v_2 - v_1)x(1 - x) = \Delta v x(1 - x),$$

b) If the Wrightian fitness of wild type and mutant are $w_1 = w$ and $w_2 = w(1 + s)$ what is the relationship between Δv and s ?

c) If an arbitrary constant c is added to each fitness (e.g. I decide to measure fitness of wild type and mutant relative to fitness of another (very closely related) organism or strain), does this change the ODE for x? If not, why not?

Question 3   Fisher's fundamental theorem of natural selection

In the lectures we derived the equations for the allele frequency in the next generation

$$x_{t+1} = \frac{(1+s)x_t}{\overline{w}},$$

where $\overline{w} = 1 + sx_t$ is the mean (Wrightian) fitness of the population.

a) Show that $\delta x_t = x_{t+1} - x_t$ is given by

$$\delta x_t = \frac{sx_t(1-x_t)}{\overline{w}}$$

b) Show that this can be expressed as

$$\delta x_t = \frac{x_t(1-x_t)}{\overline{w}}\frac{d\overline{w}}{dx} = x_t(1-x_t)\frac{d\ln\overline{w}}{dx}$$

(Hint: for last part, use chain-rule for to show equivalence between LHS and RHS)

c) This result is often called Fisher's fundamental theorem of natural selection, although Fisher's result was related to a continuous time description and that populations increase their fitness proportional to the genetic variance in fitness. Here, interpret what $\ln\overline{w}$ represents.

d) For a population that has K-alleles the frequency of the kth allele in the next generation is given by

$$x_k(t+1) = \frac{w_k(x_k(t))}{\overline{w}},$$

Where $\overline{w} = \sum_{k=1}^{K} w_k \times x_k(t)$. Show that the change in frequency of the kth allele is given by

$$\delta x_k(t) = \frac{w_k - \overline{w}}{\overline{w}} x_k(t),$$

And interpret what this means for whether the frequency of the kth allele increases or decreases in the population.

Question 4 Diploid selection equations

In the lectures the dynamics of selection was described for haploid organisms. For diploid organisms we need to consider the fitness of genotypes A_1A_1, A_1A_2, A_2A_2. This question will calculate the equivalent equations for the frequency x_{t+1} of the *allele* A_2 in the next generation in terms of the frequency in the current generation x_t.

a) If the *absolute* viability fitness of each of the genotypes A_1A_1, A_1A_2, A_2A_2 is w_{11}, w_{12}, w_{22}, respectively, in generation t, what are the zygotic frequencies of i) the A_1A_1 genotype, ii) the A_1A_2 genotype, and iii) the A_2A_2 genotype in terms of the allele frequency x_t. Check that the sum of i),ii) and ii) equal 1.

b) If N_t is the total zygotic population size, what are the number of individuals in adulthood after viability selection with i) n_{11}, with A_1A_1 genotype, ii) n_{12}, with A_1A_2 genotype, and iii) n_{22} with A_2A_2 genotype.

c) Assuming random fusion of gametes and the Hardy-Weinberg law these will also be the number of zygotes in the next generation. Using the zygotic frequencies in a) show that the total population size in next generation is $N_{t+1} = \bar{w}N_t$, where $\bar{w}$ is the mean viability fitness of zygotes in generation t.

d) Given your answer to bi) and bii) above, what is i) the total number of copies n_1 of the A_1 allele and ii) n_2 of A_2 in the population?

e) Using your answer to c) what is the total number of chromosomes $2N_{t+1}$? Express your answer in terms of x_t, using your results to b).

f) Hence, using your answer to c) & d) show the frequency x_{t+1} of the A_2 allele in the next generation in terms of the absolute fitness of genotypes and x_t is

$$x_{t+1} = \frac{w_{22}x_t^2 + w_{12}x_t(1-x_t)}{w_{11}(1-x_t)^2 + 2w_{12}x_t(1-x_t) + w_{22}x_t^2} = \frac{w_{22}x_t^2 + w_{12}x_t(1-x_t)}{\bar{w}}$$

g) Now let the absolute fitness of each genotype be decomposed into a density-dependent part $w(N)$ and a relative fitness for each genotype:

$$w_{11} = w(N) \times 1; \; w_{12} = w(N) \times (1+hs), ; w_{22} = w(N) \times (1+s),$$

where s is the selective advantage of the homozygotes A_2A_2, hs is the selective advantage of the heterozygotes A_1A_2, and where h is the dominance coefficient of heterozygotes. Show that

$$x_{t+1} = \frac{(1+s)x_t^2 + (1+hs)x_t(1-x_t)}{1 + 2hsx_t(1-x_t) + sx_t^2}$$

h) Show that

$$\delta x_t = x_{t+1} - x_t = \frac{s(h + (1-2h)x)}{1 + 2hsx_t(1-x_t) + sx_t^2}x_t(1-x_t)$$

And that if $s \ll 1$

$$\lim_{\delta t \to 0}\left\{\frac{\delta x_t}{\delta t}\right\} = \frac{dx}{dt} \approx s(h + (1-2h)x)x(1-x)$$

This is the ODE continuous time description of the dynamics of selection for diploid populations.

Question 5   Diploid mutation-selection balance in deterministic population genetics

An ODE that describes deleterious selection on a mutant allele and mutation between mutant and wild type alleles for diploid individuals with dominance coefficient h is

$$\frac{dx}{dt} = -s_d(h + (1 - 2h)x)x(1 - x) + \mu(1 - 2x)$$

where $x(t)$ is the frequency of the mutant allele that has selective advantage $s = -s_d$ ($s_d > 0$) and μ the mutation rate and we assume there is incomplete dominance ($0 \leq h \leq 1$).

a) For which values of s_d is the ODE above valid as a good description for discrete dynamics

b) If h is not small, $\mu \ll s_d$ and $x \ll 1$, show that the approximate ODE is

$$\frac{dx}{dt} = -hs_dx + \mu,$$

    and that the fixed point is

$$x^* \approx \frac{\mu}{hs_d}.$$

c) However, this is not correct when $h \to 0$. In this case the ODE becomes

$$\frac{dx}{dt} \approx -s_d(h + x)x(1 - x) + \mu(1 - 2x).$$

    Again, make the approximation that $x \ll 1$ to show that the fixed point is given by

$$x^* = \frac{-(hs_d + 2\mu) \pm \sqrt{(hs_d + 2\mu)^2 + 4\mu s_d}}{2s_d}$$

$$= -\left(\frac{h}{2} + \frac{\mu}{s_d}\right) + \sqrt{\left(\frac{h}{2} + \frac{\mu}{s_d}\right)^2 + \frac{\mu}{s_d}}.$$

    Why do we take the positive square root? Why can we not ignore the term $-2\mu x$ in this calculation, when we could in a)?

d) Plot x^* vs h (x^* on a log scale) for $\frac{\mu}{s_d} = 10^{-3}$ and $0 \leq h < 1$ from b) and from c). What do you notice?

e) Show from the answer to c) that

$$\lim_{h \to 0} x^* \approx \sqrt{\frac{\mu}{s_d}},$$

    & confirm that the y-intercept of your plot corresponds to this value. (Hint: $\sqrt{x} \gg x$ for $x \ll 1$). The mutational load is a measure of how far the population is from optimum. Here if there were no mutations, the wildtype would fix and the mutational load would be $\mathcal{L} = 0$. Assuming that the optimum has fitness 1, then in general the mutational load is given by

$$\mathcal{L} = 1 - \bar{w},$$

    where $\bar{w}$ is the mean fitness:

f) Show that the mean fitness is

$$\bar{w} = 1 - 2hs_dx(1 - x) - s_dx^2$$

    And so, the mutational load is

$$\mathcal{L} = 2hs_dx(1 - x) + s_dx^2.$$

g) For finite (non-zero) h, show that the mutational load is

$$\mathcal{L} \approx 2\mu$$

h) For the case that $h \to 0$, show that the mutational load is

$$\mathcal{L} \approx \mu$$

i) Explain why the mutational load does not depend on the deleteriousness of the allele s_d.