

# Population Genetics: Assignment 1

University of Cambridge

Henrik Åhl

March 9, 2017

## Abstract

This is an abstract abstract.

## Preface

This is an assignment report in connection to the *Population Genetics* module in the Computational Biology course at the University of Cambridge, Lent term 2017. All related code is as of March 9, 2017 available through a Github repository by contacting [hpa22@cam.ac.uk](mailto:hpa22@cam.ac.uk).

## Exercises

### 1 – Measurement of variance

A Table 1: Solution to exercise 1a

Selected	¬Selected	Total [%]
0.18	0.12	0.14
0.49	0.45	0.47
0.32	0.44	0.39

B The heterozygosity is the frequency of the middle row in table 1.

C We calculate  $F_{ST} = 1 - \frac{2p_S q_S}{2p_T q_T} = 0.67$ . A low value in this would correspond to a situation where no difference in heterozygosity is prevalent between the subpopulations. In contrast, a high value would mean that the populations are completely segregated, with the respective alleles in each subpopulation being fixed.

### 2 – Modelling fitness in a diploid system

A The Hardy-Weinberg proportions simply correspond to the combinatorial probability of achieving a certain setup of alleles. Like before, it is

thus simply  $p * 2$  for genotype AA,  $2p(1 - p)$  for Aa and aA, and  $(1 - p)^2$  for aa.

B The mean fitness is given by evaluating the different fractions which are affiliated with what fitness. Retaining the algebraic form of our fitness values before calculating, we get the following expression and subsequent result:

$$\begin{aligned}\bar{f} &= (0.9a + 0.1)p^2 + 2(0.9b + 0.1)p(1 - p) \\ &\quad + c(1 - p)^2 = \\ &= 0.84\end{aligned}$$

C Reducing our expression above further, we can from here choose to differentiate with respect to the derivative. There is only extremum, which must be the maximum as the second derivative evaluates to  $< 0$ .

$$\begin{aligned}\bar{f} &= 0.82p^2 + 1.82p(1 - p) + 0.7(1 - p)^2 \\ \frac{d\bar{f}}{dp} &= 1.64p + 1.82(1 - 2p) - 1.4(1 - p) \stackrel{!}{=} 0 \\ p^* &= 0.7 \\ \bar{f}_{p^*} &= 0.85\end{aligned}$$

D We transform our problem accordingly:

Table 2: Solution to exercise 2d

Genotype	Required form	Transformed form
AA	$1 + 2\sigma$	1.17
Aa	$1 + 2h\sigma$	$1.29b + 0.14$
aa	1	1

With our transformed values, solving for the unknowns gives us  $\sigma = 0.086$  and  $h = 7.5b - 5$ .

We can thereafter investigate for which values in our equation governing the rate of change is positive. The r.o.c. is given by

$$\frac{dq}{dt} = 2\sigma p(1-p)(p + h(1-2p))$$

where  $p \in [0, 1]$ . Only the last factor will therefore affect the sign of the derivative. Inserting our values reduces the informative part to

$$p - (7.5b - 5)(1 - 2p) \stackrel{!}{\geq} 0$$

where we now want to find the values fulfilling this. Solving for our parameters gives us that this occurs in the three regions

$$\begin{aligned} b &\in [\frac{2}{3}, 0.8] \\ b > 0.8, p &\leq \frac{5(3b-2)}{2(15b-11)} \\ b &\leq \frac{2}{3}, p \geq \frac{5(3b-2)}{2(15b-11)} \end{aligned}$$

where of course the typical intervals have to be fulfilled.

### 3 – Dynamics of allele frequency change

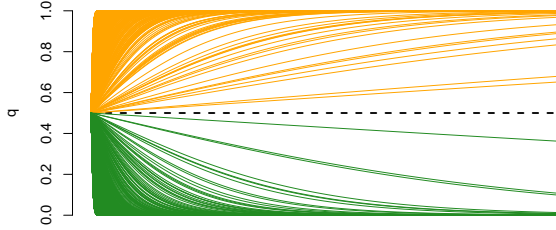


Figure 1: Stuff

A  $\mu, \sigma$  and  $N$  denote the mutation rate, the selection rate for a given allele, and the population size.

B Under selective pressure, the allele is bound to either fixate or to simply die out. Whichever effect happens depends on the sign of  $\sigma$ , i.e. for  $\sigma > 0$ , the allele will eventually fixate with  $q_i^1 = 1$ . In the contrasting case, the other allele will do the same. We can see this in fig. 1, where we have separated randomly drawn positive values of  $\sigma$  above in orange, and correspondingly all negative values in green.

C With no selective pressure, and mutations frequently producing either allele, as well as under no genetic drift, we will get the effect outlined in fig. 2, namely that the population will trend towards an equilibrium with an even split between the fractions. How fast this effect happens depends inherently on the mutation rate, which is signified in the figure by colour.

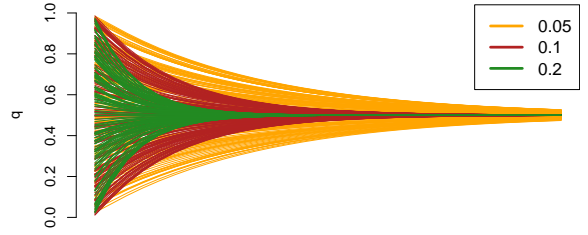


Figure 2: Stuff

D Figure 3 shows 500 drift-driven simulations initialised at  $q = 0.2$ . As we can see, the population on the large spreads out uniformly due to no pressure in going towards either end. We can argue about how likely it is on the grand scale that a single simulation will reach an allele fraction of  $q = 1$  by considering Kimura's equation first. Since the overall expression reduces to  $\frac{\partial p}{\partial t} = 0$  for  $x = 1$ , the probability for all initial conditions is some constant, which must depend on the initial condition combinatorically, as there are a different number of options to go in either direction from the beginning. The variance will only affect the rate at which we reach a certain point, so we can disregard this from our

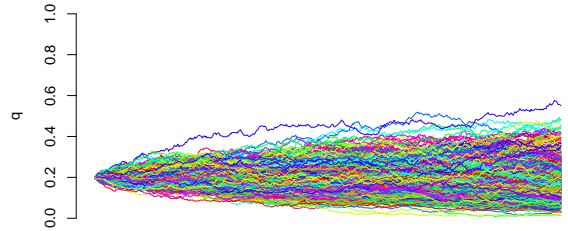


Figure 3: Stuff

## 4 – Time-dependent selection

A We separate our expression into two parts, and use induction to reason our way to the final answer. In particular, we have

$$\begin{aligned}
 x(t') &= \frac{x_0 e^{\sigma_1 t'}}{1 - x_0 + x_0 e^{\sigma_1 t'}} \\
 x(t > t') &= \frac{x_{t'} e^{\sigma_2 t}}{1 x_{t'} + x_{t'} e^{\sigma_2 t}} = \\
 &= \frac{x_0 e^{\sigma_2 t + \sigma_1 t'}}{(1 - x_0 + x_0 e^{\sigma_1 t'})} \times \\
 &\quad \times \frac{1}{\left(1 - \frac{x_0 e^{\sigma_1 t'} + x_0 e^{\sigma_1 t' + \sigma_2 t}}{1 - x_0 + x_0 e^{\sigma_1 t'}}\right)} = \\
 &= \frac{x_0 e^{\sigma_1 t' + \sigma_2 t}}{1 - x_0 + x_0 e^{\sigma_1 t' + \sigma_2 t}}
 \end{aligned}$$

for some arbitrary intervals where our  $\sigma$ 's are separable. Since we can repeat this process for any number of intervals, our corresponding exponent will equal a sum over the  $\sigma - t$  products, which in the limit of our timesteps trending towards zero is equivalent to our sought-after integral.

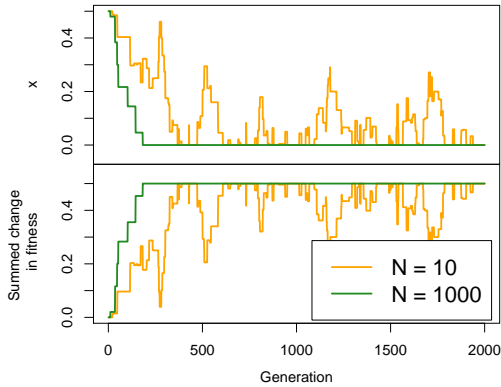


Figure 4: Stuff

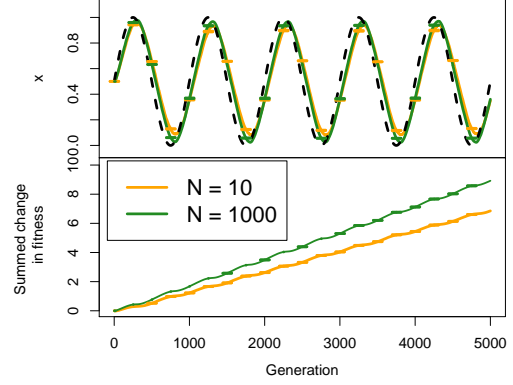


Figure 5: Stuff

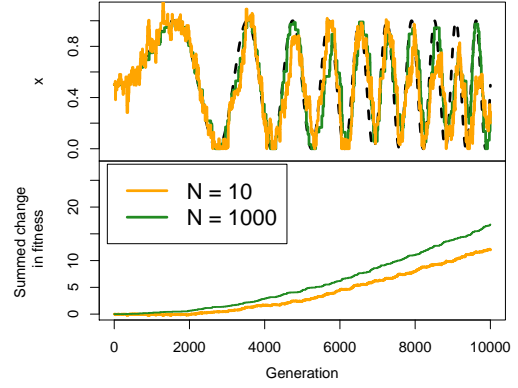


Figure 6: Stuff

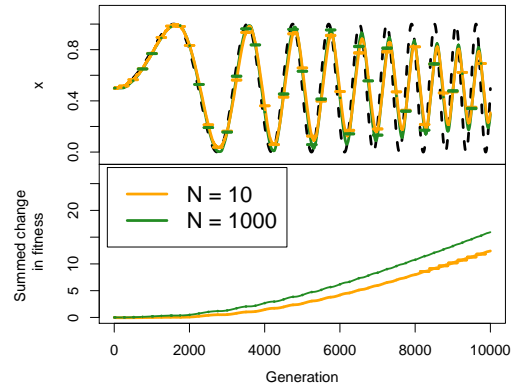


Figure 7: Stuff

Biological systems: Some populations have a harder time adapting.