Quantitative Genetics Simulation Demo

Introduction

The complexity of quantitative genetics often requires intensive computation tools for analysis. Researchers can also use computers to run simulations of breeding programs. These can be especially informative for hypothesis testing or for decision-making in breeding. We will use simulations to better-understand the concepts of population genetics, genetic variance, and response to selection.

Setup Instructions

If you have not yet installed the `shiny` package, do so using the following command:

```
install.packages("shiny")
```

Initiate the Simulation

1. First, load the `shiny` package by running the following command:

```
library(shiny)
```

2. Next, start the simulation by running this command:

```
runGitHub("neyhartj/qgdemo")
```

3. A window will pop up in your internet browser (or in a preview window if you are using R Studio). You are ready to go!

Please read the instructions provided in the introduction, and then use the following questions to guide your interaction with the simulation.

Randomly-Mating Populations

Breeding programs rely on **breeding populations**, which are defined by **genotype frequencies** and **allele frequencies**. These frequencies can be predicted if populations are in Hardy-Weinberg Equilibrium.

For the first simulation, we will look at the change in allele frequency at a single gene in a randomly-mating breeding population.

Problem 1

Without changing parameters, click the "Simulate" button to run an iteration of the simulation. You will notice that the allele frequency is changing over generations. Are the populations in Hardy-Weinberg Equilibrium? Inspect the different parameters that you can change (i.e. allele frequency, population size, number of generations). What is the most obvious violation of the assumptions of Hardy-Weinberg Equilibrium?

The populations are NOT in Hardy-Weinberg Equilibrium (HWE). A fundamental outcome of HWE is that allele frequencies do not change over generations, and since the allele frequency is clearly changing in this simulation, the populations are not in HWE.

The assumptions of HWE are: 1) no mutation, 2) no migration, 3) infinite population size, 4) no selection, and 5) random mating. The most glaring violation of these assumptions in this simulation is the lack of *infinite population size*.

Problem 2

Even though our populations are randomly-mating, allele frequencies can change over time simply by random chance, or **genetic drift**. Now, increase the breeding population size (try something large like **1,000** or **1,500** individuals) and re-run the simulation. What can you conclude about the effect of population size on the level of genetic drift (i.e. the random changes in allele frequency)?

As the population size increases, the level of genetic drift decreases.

Problem 3

Reset the population size to around **100** individuals and re-run the simulation. In any of the populations (defined in different colors), did the alleles reach **fixation** (i.e. p = 0 or 1)? If not, you may re-run the simulation until this occurs.

Variation at a gene is lost when alleles reach fixation. This can happen simply as a result of **genetic drift**. Now, with the population size at around **100** individuals, change the starting allele frequency to be close to 0 (i.e. < 0.1) or close to 1 (i.e. > 0.9). Re-run the simulation. What can you conclude about the effect of starting allele frequency on the likelihood that an allele reaches fixation?

The likelihood that an allele reaches fixation increases as the allele approaches a very high or very low frequency. In fact, the probability that an allele reaches fixation in a randomly mating population is equal to its frequency.

Problem 4

Keep the starting allele frequency close to 0 or 1. Adjust the population size to a large number of individuals. What can you conclude about the impact of starting allele frequency and population size on the likelihood that an allele reaches fixation?

As population size increases, alleles are less likely to reach fixation. This is because at a higher population size, the population behaves more like one in Hardy-Weinberg Equilibrium (HWE). Since allele frequencies do not change in HWE, an allele cannot become fixed.

Problem 5

Allele frequencies can change due to random chance, even in the absence of selection. Speculate on the importance to breeding of losing variation at a gene due to allele fixation.

Variation that is lost to fixation due to genetic drift cannot be selected upon by a breeder. If, for instance, alleles that confer unfavorable grain yield become fixed in the population due to genetic drift, the breeder will no longer be able to select for the favorable alleles. Thus breeding progress will be reduced
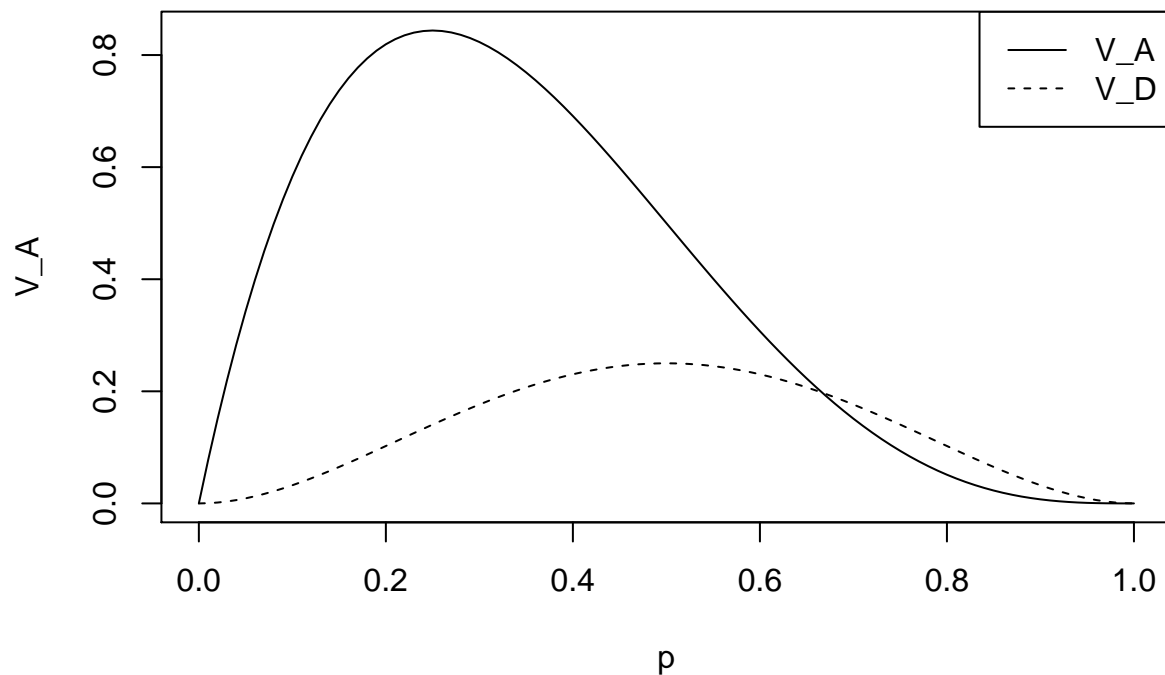
Genetic Variance

Move on to the Genetic Variance tab. This simulation will assess the impact of **allele frequency** and **coded genotypic values** on the level of genetic variance at a single gene.

Problem 1

Adjust the starting allele frequency (the simulation will automatically update). What is the effect of starting allele frequency on the ratio of additive variance (V_A) to dominance variance (V_D)? What happens when the allele is **fixed** (i.e. p = 0 or 1)?

At a low starting frequency of the positive allele, the proportion of genetic variance due to additive genetic variance is much larger than the proportion due to dominance genetic variance. This trend reverses as the allele frequency increases. To visualize this, see below:

When the allele frequency is 1 or 0, there is no genetic variance.

Problem 2

Set the starting allele frequency back to around **0.5**. Adjust the additive effect (a) and level of dominance (d). What can you conclude about impact of the ratio of a to d on the ratio of V_A to V_D?

A higher ratio of a to d gives a higher ratio of V_A to V_D.

Problem 3

Set a to **0** and d to a value larger than 0 (say, **5**). Play with the starting allele frequency. Is it possible to have $V_A > 0$ if a is equal to 0? Why is this?

Yes this is possible. The formula for V_A in a randomly-mating population is $V_A = 2pq[a+d(q-p)]^2$. Therefore even at $a = 0$, the dominance d contributes to V_A, as long as the allele frequency is not 0.5.

Response to Selection

Move on to the Response to Selection tab. The objective of a breeding program is to make selections to develop superior cultivars of plants. When breeding for a quantitative trait, such as yield, plant breeders aim to increase the **mean** of the population. The change in the population mean for a trait from one generation to the next is called the **response to selection**.

In this simulation, we will explore how different parameters impact the response to selection and other measurements for a trait influenced by many genes.

Without changing parameters, click the "Simulate" button to run an iteration of the simulation. The top graph shows the response to selection curve over generations. The lower-left graph shows the change in **genetic variance** across all of the genes impacting the trait. The lower-right graph shows the change in **allele frequency** for the value of p at all of the genes influencing the trait.

Problem 1

In this simulation, p measures the frequency of the allele with **positive additive effect** a. Change the starting allele frequency from **0.5** to close to 1 (i.e. > 0.9) and re-run the simulation. What do you notice about the steepness of the **response to selection** curve and the **genetic variance** curve?

At higher levels of p, the response to selection curve is much steeper than at low values of p. This is because the favorable alleles are at a higher frequency with larger values of p, and therfore the selection *limit* is reached in fewer generations. Correspondingly, the genetic variance drops much more rapidly when the favorable alleles are in higher frequency.

Problem 2

Change the starting allele frequency to close to 0 (i.e. < 0.2) and re-run the simulation. What do you notice about the steepness of the **response to selection** curve and the **genetic variance** curve?

In this case, lower values of p mean the favorable alleles are in lower in frequency, and thus it takes more time for the selection *limit* to be reached. The genetic variance curve is interesting because genetic variance increases at first, and then decreases. This is because the favorable alleles are becoming more frequent (remember additive genetic variance is highest at p = q = 0.5), but then as the favorable alleles become even more frequent, the genetic variance decreases.

Problem 3

Inspect the change in **allele frequency**. Each line tracks the value of p at a different gene. Thicker lines indicate alleles with larger value of a and thinner lines indicate alleles with smaller values of a. You will notice that some alleles reach fixation. What can you conclude about the relationship between the value of a and the number of generations required for the allele to reach **fixation**?

Alleles with a larger effect tend to reach fixation more rapidly that alleles with a smaller effect. Those alleles with larger effect contribute more to the genotypic value (and thus the phenotypic value upon which the breeder makes selections) and are therefore respond (i.e. by changing frequency) more readily to selection.

Problem 4

Reset the simulation by refreshing the browser. Play with the value of **heritability** (h^2). What do you notice about the effect of heritability on the steepness of the **response to selection** curve?

A highly heritable trait will lead to a steeper response to selection curve than a lowly heritable one. This is because as heritability increases, the phenotypic values more closely reflect the true genotypic value. When this occurs, selection acts more accurately on the genes that contribute to the trait, leading to a higher response.

Problem 5

Adjust the heritability tp **0.5** and play with the value of the **selection intensity**. What do you notice about the effect of selection intensity on the steepness of the **response to selection** curve?

Selecting more intensely leads to a higher response to selection, and thus a steeper curve.

Problem 6

Return the selection intensity to 0.1 and increase the **number of breeding generations**. What do you notice about the trend in **response to selection**? What is the relationship between the **response to selection** and the **genetic variance**? Speculate on the reason for this relationship. **Hint:** observe the change in allele frequencies for each gene.

The increase in the breeding generation is simply here to show how when the genotypic value increases as a result of selection, the genetic variance decreases. This is because the favorable alleles are becoming fixed. When alleles become fixed (i.e. p = 0 or p = 1), the genetic variance decreases. This is not necessarily a bad thing if selection has been effective in increasing the mean of the population. You will notice that if you set the heritability to 0 that genetic variance decreases as a result of genetic drift, however the mean genotypic value in the population does not improve much.

Optional Problem

Continue to adjust parameters and re-run simulations. What other observations and conclusions can you draw about the relationship between these parameters and the **response to selection**?