

# Equilibrium, selection and genetic drift in evolution

Abhishek Dasgupta

September 5, 2007

## Abstract

This report is on a exploratory laboratory simulation of the Hardy-Weinberg Equilibrium using coloured balls. A basic analysis of the data is done, in which we observe the relation between the population size and allele frequency over generations. We also show the trends in selection and genetic drift in our simulated population.

**keywords** Hardy-Weinberg equilibrium, selection, genetic drift

## Introduction

The Hardy-Weinberg equilibrium is a mathematical model which predicts that allele frequencies will be in equilibrium, provided no **mutation, selection and migration** takes place. Thus the Hardy-Weinberg equilibrium is kind of a *null hypothesis* against which we can measure the variation from equilibrium. In this experiment we shall describe a model to simulate Hardy-Weinberg equilibrium, visualise the results and try to understand the reasons behind the variation from equilibrium.

## The Hardy-Weinberg Model

What the model predicts: A certain population will maintain the same allele frequency throughout generations under certain assumptions.

Those assumptions are:

- The organism must be sexually reproducing.
- It is diploid
- Random mating
- Infinite population: This factor is especially important to note, as by varying this factor in our simulation, we see interesting variations from the equilibrium.
- No selection, mutation or migration.

Let's consider a locus which has two alleles (A and a, with A denoting the dominant allele) with allele frequencies  $p$  and  $q$ . After randomly selecting alleles and pairing we can easily see that the expected genotypic frequency of AA is  $p^2$ , of Aa is  $2pq$ , and of aa is  $q^2$ . Now the expected allele frequencies will become:

- for A:  $\frac{p^2 + pq}{p^2 + 2pq + q^2}$ , which after simplifying becomes  $p$ .
- for a:  $\frac{q^2 + pq}{p^2 + 2pq + q^2}$ , which becomes  $q$ .

(as  $p + q = 1$ ) Thus the allele frequency and genotypic frequency is *at equilibrium*.

## Our Experiment

**Materials Used:** Blue and orange balls as alleles denoted by B and O respectively.

We started with an initial allele frequency of 0.5 for both blue and orange balls. We created a gamete pool of 500 alleles each and randomly picked two balls creating a genotype. The frequencies of the respective genotypes (blue-blue, blue-orange, orange-orange) were noted down. We then assumed that each “individual” in our model had equal fecundity, and produced 10 gametes. For the heterozygotes we divided by two after multiplying by ten, as only half the gametes produced by a heterozygote are of B or O type. The gamete pool for the next generation was thus created and the process was repeated.

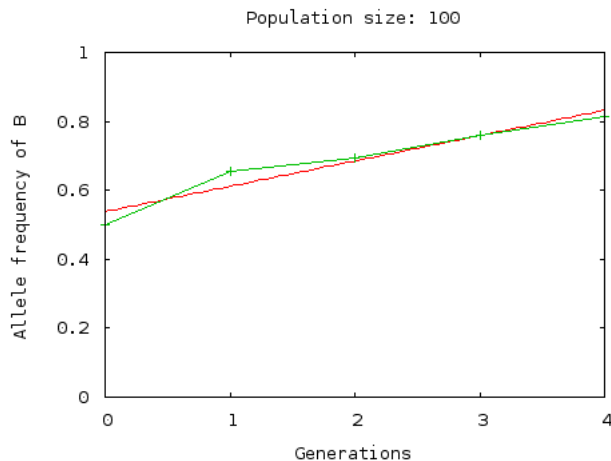
We went through this process for five to six generations, and observed the change in allele and genotype frequency. We carried out this simulation for populations of 100, 50, 40, 20, 10 and 5 individuals. The results, as we shall see, illustrate important evolutionary factors.

All trials were duplicated in two groups with nearly same conditions.

## Results and Analysis

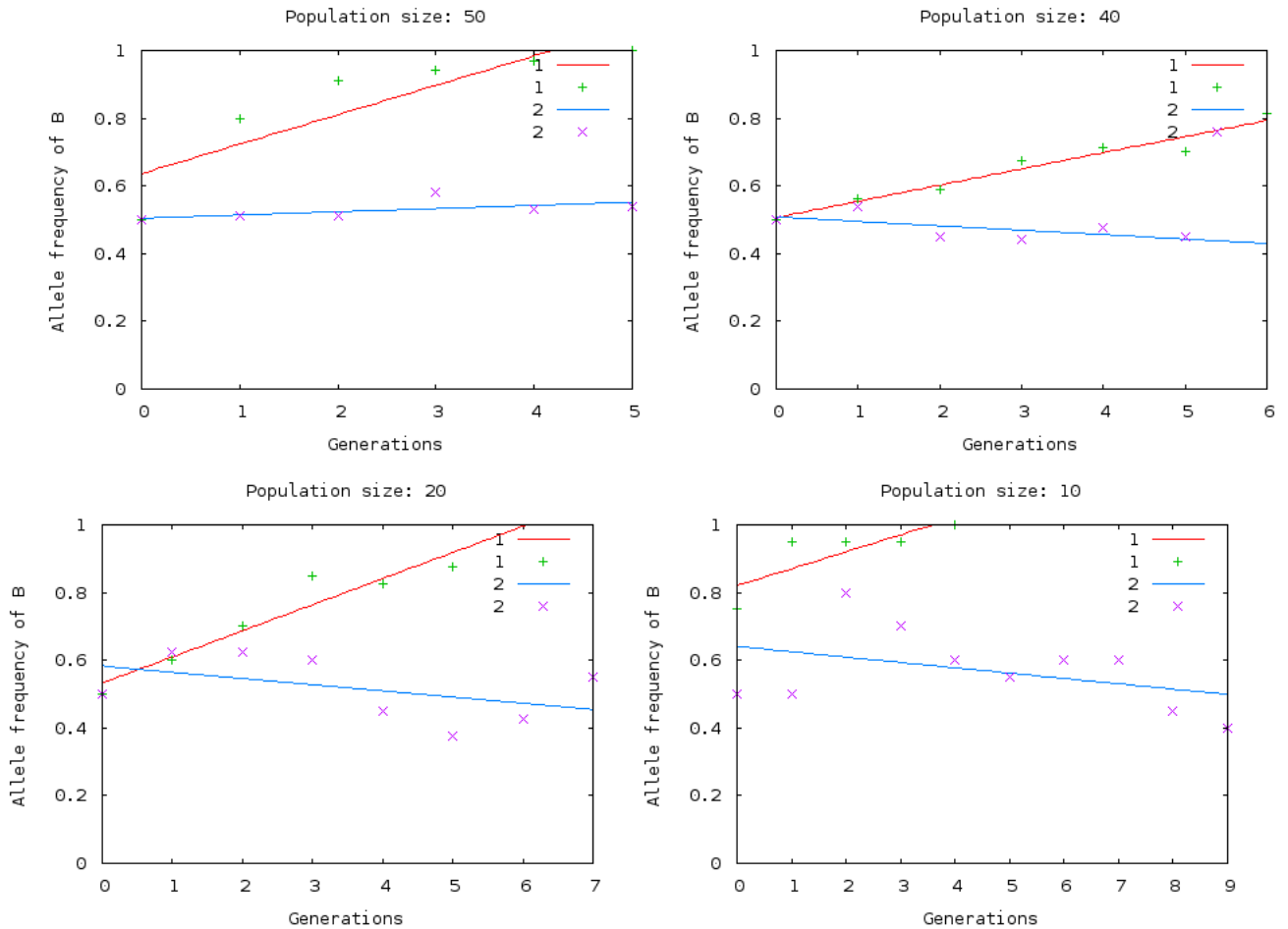
The information obtained is represented graphically. The experiment was done over two days. On the first, we recorded data for population size ( $N$ ) = 100. The other population sizes were done on the second day.

**First day:**

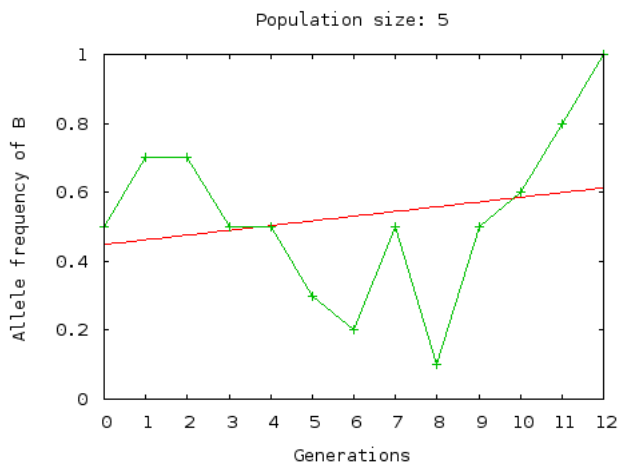


This graph clearly shows an increasing trend for the blues over the generations. Considering the conditions for maintenance of Hardy-Weinberg equilibrium, we see that we’ve no migration or mutation. So the only possible cause for this is **selection**. A possible cause of selection is the varying size of the blue and orange balls, and an inherent bias while picking the genotypes for the next generation towards picking the blue ones due to their larger size. We can test this hypothesis by blindly selecting larger balls and seeing that the selection pressure is reduced, and in some cases, even reversed.

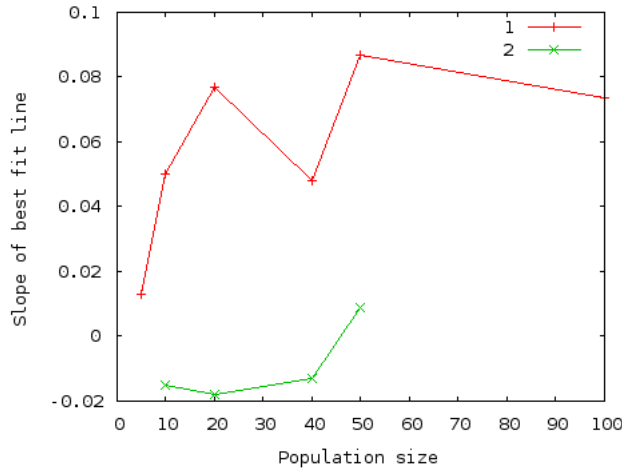
**Second day:** We varied the population from 50 down to 5, and did a best fit line to indicate the direction of selection (if any) acting on the population.



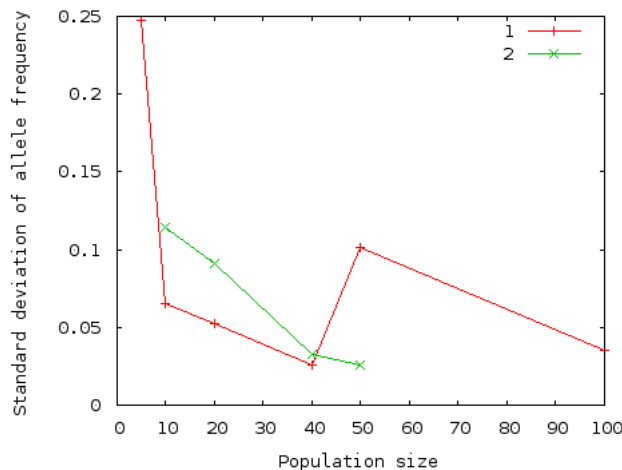
All these graphs show a similar trend. Two sets of observations are plotted for the two groups marked **1** and **2** in the graphs. Clearly group 1 still shows some effect of selection right from population size ( $N$ ) = 50 to  $N$  = 10. However group 2 data shows a marked deviation from what we would have expected considering the selection pressure. This fact can be explained by the *sampling error* that is caused as population size reduces. In lower population sizes, allele frequency estimates from the equilibrium become farther from the actual allele frequencies due to this error. This phenomenon is known as random genetic drift. In fact though group 2's data gives an illusion of equilibrium, actually it is random genetic drift which is responsible for keeping their allele frequency close to the initial value of 0.5 (Random changes tend to cancel each other out). We have hypothesised that the differing size of the bowls in which we constructed our gamete pools is the cause for the variation in results between the two groups. As our bowl was smaller, selection was prominent, while in a bowl with larger base, there would be more distance between the orange and blue balls and thus while picking there would be lesser bias to picking the blue ball.



This graph illustrates an extreme case of random genetic drift with a population size of only 5. The results were not duplicated in the other group, as the randomness of the data means that environmental effects would be washed out. The allele frequency varies erratically, however the best fit line remains nearly at 0.5 with a small positive slope. We can consider the **slope** of the best fit line to be an indicator of how much selection contributes to evolution. This consideration is justified as we see the slope of the group 2 data (where random genetic drift overrode selection) is consistently low. The fact that contribution of selection decreases with population can be seen by the following graph:



Also we can get an idea of the genetic drift by seeing the standard deviation of the allele frequency [2]. This is reasonable as more genetic drift will cause the allele frequency to fluctuate about rapidly, thus increasing the standard deviation. Seeing the graph, we can clearly see the marked increase in standard deviation and thus the genetic drift for  $N = 5$ .



## Conclusion and Further Work

This experiment was an illustrative one because it showed two important aspects of evolution—selection and genetic drift. We started off with the intent to simulate Hardy-Weinberg equilibrium, however our data clearly deviated from it. In analysing the results, we saw the role of selection and genetic drift in small populations. Analysis on reduced populations is important in endangered species research where often better alleles get extinct due to random genetic drift in their small populations.

Further work on this experiment could be a better numerical analysis of the data, for example using  $\chi^2$  analysis. We could also extend this simulation to introduce more real-world aspects like migration, mutation and fitness variation. I did perform a basic fitness analysis assigning fitness values of  $a$  to OO,  $a + b$  to BO, and  $a + 2b$  to BB to glean information on fitness (assuming additive variation). However, the fitness values varied quite a bit, though there was a trend of decreasing fitness for heterozygotes as expected.

## Acknowledgements

My heartfelt thanks to Dr. N G Prasad, who helped us greatly in the execution of this experiment, and clarified our conceptual doubts. I also thank my groupmates with whom I worked on this experiment.

## References

1. Wikipedia: Hardy-Weinberg Equilibrium, [http://en.wikipedia.org/wiki/Hardy-Weinberg\\_principle](http://en.wikipedia.org/wiki/Hardy-Weinberg_principle)
2. Inbreeding and Genetic drift, <http://www.ucl.ac.uk/~ucbhdjm/courses/b242/InbrDrift/InbrDrift.html>
3. Introduction to Population Biology, Dick Neal.