Simulation of Hardy-Weinberg Equilibrium and Study of the Deviations from the Equilibrium conditions.

by

Sambit Bikas Pal (06MS03), Indian Institute of Science Education & Research, Kolkata.

Abstract: *In this work, the Hardy Weinberg Equilibrium in a population was simulated using an artificial population of coloured polystyrene balls. Two balls at a time were picked up randomly to create a genotype. The process was repeated several times and the allele frequency was noted in every generation.*

Keywords: *Hardy-Weinberg Equilibrium, Selection, Simulation, Genetic Drift.*

Introduction: In population genetics how the frequencies of alternative states of genes, that is alleles in populations are maintained or changed from generations to generations is of prime importance. The rediscovery of Mendelian genetics led people to speculate about its implications for the genetic structures of populations. G.H. Hardy and W. Weinberg showed that Mendelian dominance had no effect on allelic frequency, and furthermore proved that allelic frequencies would not normally change as a result of sexual reproduction.

The Hardy Weinberg Equilibrium which states that the allelic frequencies remain constant from generation to generation, holds good under some assumptions, namely:

1. The organism is sexually reproducing.
2. It has discrete generations.
3. The allele under consideration is on an autosomal chromosome.
4. The population is idealised which itself implies:
   - The population size is infinite.
   - There is random mating within the population.
   - There is no selection, no mutation and no genetic drift.
   - There is no migration or gene flow.

Derivation of Hardy- Weinberg Principle:

Let us assume that the frequencies of two alleles A & a in a population are p & q respectively.

Hence $p+q=1$

In the next generation the frequencies of AA, Aa & aa genotypes will be p^2, $2pq$ & q^2 respectively.

$p^2 + 2pq + q^2 = 1$

The frequency of A in next generation= $p^2+pq=p(p+q)=p*1=p$

The frequency of a in next generation= $q^2+pq=q(p+q)=q*1=q$

Hence the allelic frequencies remain unchanged.

The Hardy Weinberg Equilibrium is a “'neutral-equilibrium”. This means that though the allelic and genotypic frequencies do not change under random mating, if some other external agent like selection or migration, changes the allelic frequencies to new values, the genotypic frequencies shift accordingly. If no other force is applied the population will attain equilibrium at these new frequencies i.e. the population will not return to the old Equilibrium on its own.

This simulation dealt mainly with the study of deviations of a population from the Hardy Weinberg Equilibrium.

Causes of Deviation from Equilibrium in a Real Population:

The exact idealised conditions which are taken to be the fundamental assumptions of the Hardy-Weinberg Principle are hard to find in a real population. Violation of the assumptions can lead to deviation from the expected values. The degree of effect on the population is dependent on the degree of violation of the assumed conditions.

- The provision of random mating may be violated when there is inbreeding which favours homozygousity or there is assortative mating which causes an increases in homozygousity of those genes involved in the trait that is assortatively mated.
- The Hardy-Weinberg Principle expects that the population size is infinite. A large population provides a close approximation

to that assumption. But if the population size is too small, then due to sampling error random fluctuations may arise in the allelic frequencies. This is known as genetic drift, which disturbs the Hardy-Weinberg Equilibrium.

- Besides these, the equilibrium may also be disturbed by external factors like selection or mutation which tends to change the allelic frequencies.

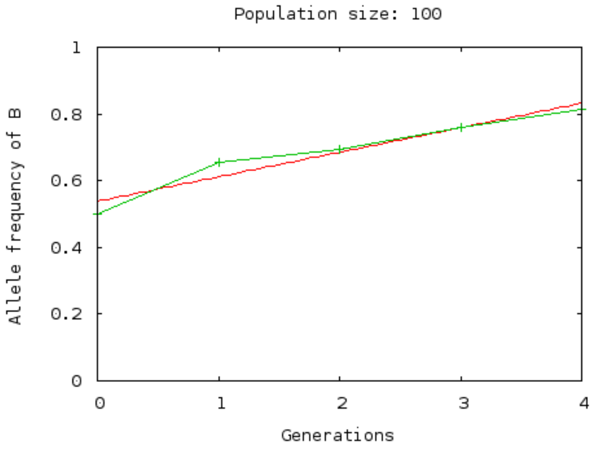
Experimental:

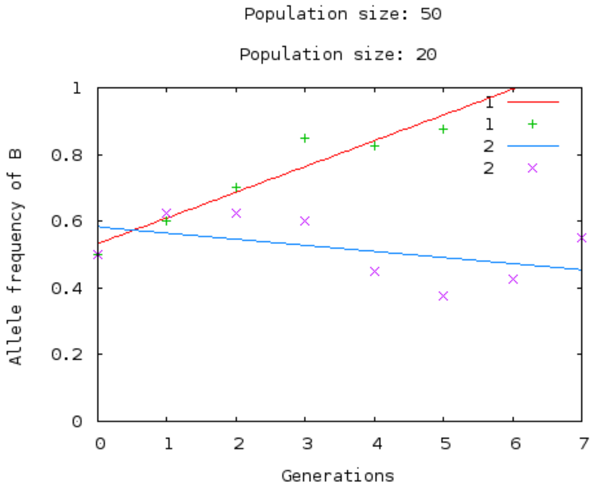
- Requirements:
  i. Polystyrene Balls of two colours.
  ii. Bowls

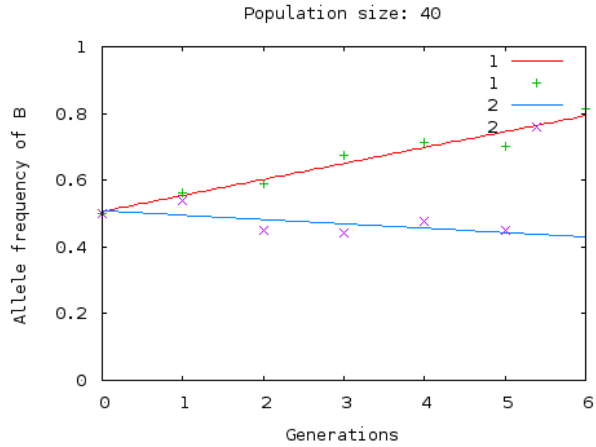
- Experimental procedure:
  i. An artificial gamete pool is created by mixing a known number of blue and orange balls.
  ii. Two balls are randomly pulled out to create a genotype.
  iii. The genotype, whether it is blue-blue, blue-orange or orange-orange is noted down.
  iv. Assuming that all the genotypes have equal reproductive ability, and arbitrarily taking the number of gametes produced by each genotype to be ten, the process was repeated N/10 times, where N is the total gamete pool size, thus maintaining the total population size in every generation.
  v. The genotypic frequencies were noted down, from which the allelic frequencies were calculated.
  vi. Initially we started off with a gamete pool of 1000 gametes, consisting of 500 blue and 500 orange balls. Two balls were picked up randomly 100 times creating a population of 100 diploids with various genotypes. The genotypic frequencies in the F_1 generation was noted down. Since each diploid organism was assumed to produce 10 gametes, the gamete pool of the next generation still had 100*10=1000 gametes, but the allele frequencies were different. 100 balls were again randomly picked up to give F_2 generation and the process was repeated for 7 generations.
  vii. The same was done for the gamete pool size of 500, 400, 200, 100 & 50 gametes.
  viii. The entire experiment was repeated twice and two data sets were obtained.

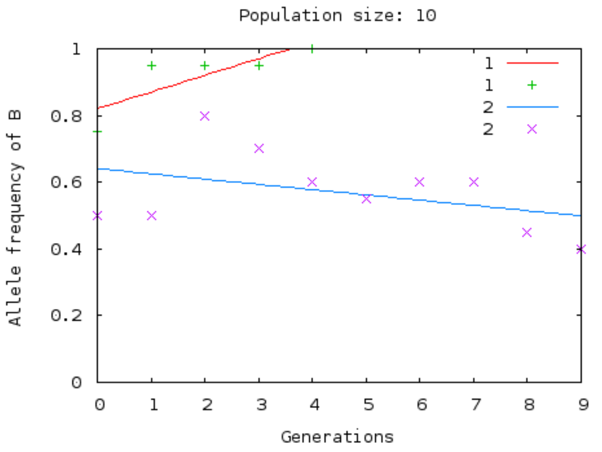
Results:

The experimental results have been depicted graphically below:









In the graphs the frequency of only the blue allele was plotted, since the plot of the frequency of the other allele will be the mirror image of the first one about y=0.5 .

The frequency curve of the blue allele was fitted to a straight line whose slope signified the rate at which the frequency of the blue allele changed.

The graphical data has been summarised & tabulated below:

Sr. No.	**Pop Size**	**Slope of the B allele freq. curve (Data Set 1)**	**Slope of the B allele freq. curve (Data Set 2)**
1	1000	0.0735	Data unavailable
2	500	0.087	0.009
3	400	0.048	-0.013
4	200	0.077	-0.018
5	100	0.05	-0.015
6	50	0.013	Data unavailable

Points to be noticed:

Some of the most glaring points noticeable in the results are:

i. The two data sets yielded quite different results.
ii. Exact equilibrium conditions were not attained in any of the cases.
iii. In case of very small gamete pool random results were obtained.

In case of the Gamete Pool size of 1000 gametes, both the Data sets gave identical curves. But there is marked deviation from equilibrium values in all the population sizes when Data set 1 is considered. Contrary to that the Data set 2 gave frequencies that deviated much less from Equilibrium values. At very low population size, however completely random results were obtained.

Interpretations:

- The random results obtained at low population size has an evidently clear explanation. The low population size leads to sampling error during mating, even though the mating may appear to be random. A consequence of this is Genetic Drift, which disrupts the equilibrium.
- The increase in the allelic frequency of the blue alleles in case of data set 1 indicates a selection favouring the blue alleles. One of the reasons for the selection of the blue balls were their larger size, which led to their more frequent picking up. Also the orange balls had a tendency of sinking down, which reduced their chances of being selected for.
- In case of the Data set 2 the selection favouring the blue was not strong enough for some reasons and the random genetic drift was over riding the weak selection, leading to attainment of almost Equilibrium values. The most probable reason for weaker selection in case of Data set 2 is that the experimenters used a larger bowl, and all the balls were equally easy to pick up

irrespective of their shape.

Conclusions:

Study of the Hardy-Weinberg equilibrium is of immense importance in population biology. The knowledge of how the equilibrium can be restored from very low allele frequencies is of great use for conservation of endangered species. As a part of continuation of this work, one can study the effects of migration, differential reproductive abilities and other effects on the equilibrium condition. Also advanced statistical tools may be used to obtain a further insight into the inner workings of population genetics.

Acknowledgements:

First of all I would like to express my heartfelt gratitude to Dr. N.G. Prasad, IISER Kolkata for his invaluable guidance in every stage of the work. Also I would like thank my colleague Abhishek Dasgupta of IISER Kolkata for his kind help in plotting the data using GNUPlot software.

References:

i. http://en.wikipedia.org/wiki/Hardy-Weinberg_principle
ii. Neal Dick. “Introduction to Population Biology”. Cambridge University Press.
iii. Hardy, G. H. (1908). "Mendelian proportions in a mixed population". *Science* **28**: 49 – 50.