Aniket Kittur
adkittur@princeton.edu
adapted from Dannie Durand
durand@cs.princeton.edu

Simulating Population Genetics

Some introductory genetics: Genes are DNA sequences whose code determines which proteins are produced, and are grouped together in chromosomes. Higher organisms have two copies of each chromosome, one from the male and one from the female; such organisms are referred to as *diploid*. Thus each organism has two, possibly different, copies of each gene (these copies of a gene are called *alleles*).

To mate, diploid cells divide to produce sex cells, such as sperm or eggs. Each sex cell is haploid; that is, it contains only one set of chromosomes from the parent instead of two. If we consider a diploid organism, say a mouse, with two possible alleles, a and A, this usually means that fifty percent of the sperm from a male Aa mouse will contain the a allele and the other fifty percent will contain the A allele. The offspring of two Aa mice would thus have a 25% chance of being aa, a 50% chance of being Aa, and a 25% chance of being AA. The combination of genes in a particular mouse is referred to as its *genotype*.

If we know the initial distribution of alleles in a population we can calculate a number of useful probabilities. Assuming we know the frequency of each allele in the population, we know the probabilities of each allele are $P(A) = p$ and $P(a) = (1 - p) = q$. From this we can calculate the probability of finding a mouse in the population with a particular genotype (AA, Aa, or aa), how these probabilities vary over time and what the steady state probabilities will be. G. H. Hardy and W. Weinberg independently solved this problem in 1908 under a set of idealized assumptions:

- The population is infinite.
- All male, female pairs are equally likely to mate.
- Alleles do not spontaneously appear or disappear from the population (i.e. no migration or mutation.)
- All alleles are equally fit.

Under these conditions, the following steady-state genotype frequencies are established:

$P(AA)$	$P(Aa)$	$P(aa)$
p^2	$2pq$	q^2

The derivation of these frequencies are explained in [1].

Your assignment: The first two conditions never hold; the remaining two conditions hold some of the time, at best. We will examine these assumptions of the model and see if and how they affect the predictions and results.

- Simulate a finite population of mice with two alleles at a single locus (e.g. A and a). Assume that mice mate randomly and that the population size is fixed; because of chance not all the mice will necessarily mate and some may mate more than once. Also assume that the generations are synchronized: all the mice in generation i are offspring of mice in generation $i-1$. It is not necessary to get fancy with implementation (e.g. don't use linked lists when arrays will suffice). Track how the frequency/probability of the alleles change over time using any relevant methods, for example histograms, means and standard deviations.

- How do the allele frequencies vary over time? Is an equilibrium condition reached? How do you recognize when (if?) equilibrium is reached and how quickly do you reach this condition, if it occurs? Does it differ from the steady state allele frequencies predicted by Hardy and Weinberg?

- How does population size affect the answers to these questions? How about the allele frequencies in the initial population? How many times should you run each simulation (i.e. each set of conditions) to have confidence in your results?

- Relax the random mating condition. Partition your population into subpopulations and only allow the mice to mate within subpopulations. What are the allele frequencies in the subpopulations? In the total population? How do they differ from the Hardy-Weinberg frequencies?

Extra Credit:

- Relax the assumption that all alleles are equally fit. Choose one of your alleles to be lethal recessive; that is, if the a allele is lethal recessive, aa mice die at birth but Aa and AA mice don't. How does this change the equilibrium? Can any starting conditions change the final equilibrium? In some inherited lethal recessive diseases, such as Huntington's chorea, alleles continue to propagate in populations in accordance with the Hardy-Weinberg predicted frequencies. How is this possible?

- Now we will model a real-life example of the forces working in evolution, using the t-haplotype condition found in mice. Mice which have two copies of a mutant t-haplotype gene (call it t) die at birth. Mice with one normal gene and one mutant gene ($+t$) seem almost exactly the same as mice with two normal genes ($++$), but have one major difference: a +/t male mouse passes t to more than 90% of his offspring, and thus passes the normal allele to less than 10%. The t-haplotype has no effect on +/t females. Can a stable equilibrium be reached with these conditions?

- Current studies suggest that there might be another force involved in the t-haplotype condition. Some researchers believe that there is sexual selection at work as well: females slightly prefer males who have two normal genes over those who have one normal and one mutant gene (of course, males with two mutant genes never get to mate). This difference is small but detectable. This type of selection is evolutionarily plausible, since it would lead to a greater number of viable offspring. Modify your simulator for the t-haplotype to determine the effects of sexual selection on t-haplotype frequencies.

For further reading and project ideas see the references section.

References

[1] J. Maynard Smith. *Problems of Biology.* Oxford University Press, Oxford, England, 1986.

[2] J. Aroesty. Tumor growth and chemotherapy: Mathematical methods, computer simulations, and experimental foundations. *Math. Biosci.*, 17:243–300, 1973.

[3] J. L. Gould. *Biological Science.* W.W. Norton, New York, 1996.

[4] * H. Gould and J. Tobochnik. *An Introduction to Computer Simulation Methods.* Addison Wesly, Menlo Park, Ca., second edition, 1996.

[5] H. W. Hethcote and J. A. Yorke. *Gonorrhea Transmission Dynamics and Control.* Springer-Verlag, New York, 1984.

[6] ** L. Edelstein-Keshet. *Mathematical models in biology.* Random House, New York, 1988.

[7] W. J. Ewens. *Mathematical Population Genetics.* Springer-Verlag, Berlin, 1979.

[8] R. E. Keen and J. D. Spain. *Computer Simulation in Biology; A Basic Introduction.* John Wiley, New York, 1992.

[9] J. F. Crow. The ultraselfish gene. *Genetics*, 118:389–391, 1988.

[10] L. M. Silver. Mouse *t* haplotypes. *Annual Review in Genetics*, 19:179–208, 1985.

[11] L. M. Silver. The peculiar journey of a selfish chromosome: Mouse *t* haplotypes and meiotic drive. *Trends in Genetics*, 9:250–254, 1993.

[12] T. W. Lyttle. Cheaters sometimes prosper: Distortion of Mendelian segregation by meiotic drive. *Trends in Genetics*, 9:205–210, 1993.

[13] J. H. Werren, U. Nur, and C.-I. Wu. Selfish genetic elements. *Trends in Ecology and Evolution*, 3:297–302, 1988.

[14] K. G. Ardlie. *The Frequency, Distribution and Maintenance of t-haplotypes in Natural Populations of Mice.* PhD thesis, Princeton University, Princeton, N.J., 1995.

[15] R. Durrett. Spatial epidemic models. In D. Mollison, editor, *Epidemic Models: Their Structure and Relation to Data.* Cambridge University Press, Cambridge, U.K., 1995.

[16] M. Mitchell and S. Forrest. Genetic algorithms and artificial life. In Christopher G. Langton, editor, *Artificial Life: An Overview.* MIT Press/Bradford Books, 1995.

[17] ** J. Maynard Smith. *Evolutionary Genetics.* Oxford University Press, Oxford, England, 1989.