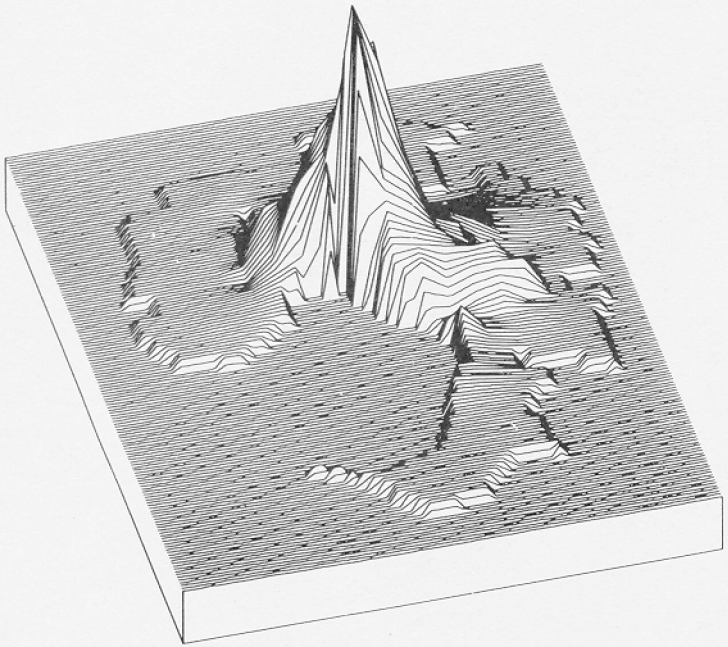
ISSN 0313-6825

PUNDIT

Volume 4 Part 5

October 1980



A SCHOOL MATHEMATICS MAGAZINE

Published by Monash University

MATHEMATICAL MODELS IN POPULATION GENETICS

Simon Tavaré, University of Utah

Population genetics is one of many biological fields in which mathematical models have played a significant role. Although such models do not claim to be precise descriptions of reality, they are nonetheless useful in assessing the roles of separate parts of the real process under study. In this article, we describe some of the simpler genetic models. We will indicate briefly the methods that may be used to analyse them, and the results that follow.

The study of theoretical population genetics attempts to quantify how the genetic constitution of a population changes with time. We will suppose that each individual in the population under study is one of three possible genetic types, denoted by AA, Aa and aa. We call the classifications AA, Aa, aa *genotypes*. Any individual inherits one *gene* (either A or a) from each of his parents. For example, if one parent is AA and the other Aa, then the offspring are AA or Aa, each with chance $\frac{1}{2}$. If the population is a type of sweet pea, and A is the gene for red flowers, a the gene for white flowers, then AA flowers are red, aa white and Aa are pink. Further details of elementary genetics are provided in reference [1] below.

We say that our population exhibits (genetic) *variation* if both the genes A and a are present in that population. This genetic diversity should allow the population to adapt more readily to its surroundings.

THE HARDY-WEINBERG LAW

The simplest result about the genetic makeup of the population is provided by the celebrated Hardy-Weinberg Law [3,4]. For simplicity we will assume that generations are discrete: at each time point all individuals die, and are replaced by new individuals at the next time point. Now suppose that the three genotypes AA, Aa, aa are represented in proportions D, $2H$, R respectively ($D + 2H + R = 1$). The fraction of A genes is denoted by p, and is given by $p = D + \frac{1}{2}.(2H) = D + H$. To compute the proportions D', $2H'$, R', p' in the next generation, we use the mating table opposite.

Mating table

mating type	proportion	fraction of offspring of type AA	Aa	aa
$AA \times AA$	D^2	1	0	0
$AA \times Aa$	$4DH$	$\frac{1}{2}$	$\frac{1}{2}$	0
$AA \times aa$	$4DR$	0	1	0
$Aa \times Aa$	$4H^2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$Aa \times aa$	$4HR$	0	$\frac{1}{2}$	$\frac{1}{2}$
$aa \times aa$	R^2	0	0	1

The proportion D' of AA genotypes is given by

$$D' = D^2.1 + 4DH.\tfrac{1}{2} + 4H^2.\tfrac{1}{4} = (D + H)^2 = p^2$$

while $2H' = 2p(1 - p)$, $R' = (1 - p)^2$. Also note that $p' = D' + H' = p$. Repeating the previous argument in the next generation shows that $D'' = p^2$, $2H'' = 2p(1 - p)$, $R'' = (1 - p)^2$, and the same proportions are maintained in all subsequent generations.

This result is extremely important. Assuming there are no external pressures (such as are caused by one genotype being more successful in producing offspring than others) acting on the population, genetic variation is maintained. Notice that the fraction, p, of A genes is constant in *all* generations.

We now turn to the 'external pressures' mentioned above. We refer to the different survival or mating success of the three genotypes as *selection*. How does selection affect genetic variation?

THE ROLE OF SELECTION

We will introduce three positive numbers w_{AA}, w_{Aa}, w_{aa} which are called the relative *fitnesses* of the genotypes. They are used to model the effect of different viability and fertility among the genotypes. If, for example, AA genotypes are more successful (at reproducing or surviving) than aa genotypes, then w_{AA} is larger than w_{aa}. Let p_n be the proportion of A genes in generation n, and let $q_n = 1 - p_n$, $n = 0,1,2,\ldots$. A mating table similar to the previous one can be used to show that p_{n+1} is related to p_n by the formula

$$p_{n+1} = \frac{p_n(w_{AA}p_n + w_{Aa}q_n)}{p_n^2 w_{AA} + 2p_n q_n w_{Aa} + q_n^2 w_{aa}}, \quad n = 0,1,2,\ldots . \qquad (*)$$

Notice that if $w_{AA} = w_{Aa} = w_{aa}$, then $p_{n+1} = p_n = \ldots = p_0$, which is the Hardy-Weinberg case we have already encountered.

Equations like (*) are called recurrence relations, and they play an important part in mathematical modelling. From our point of view the interesting question to ask is: What happens to p_n after a large number of generations? In principle, this is not a simple question to answer, but we will illustrate what happens in some specific cases.

Case (a) $w_{AA} = 0$, $w_{Aa} = w_{aa} = 1$.

This is a model in which *AA*-individuals cannot reproduce. Examination of (*) in this special case shows that

$$p_n = p_{n-1}/(1 + p_{n-1}), \quad n = 1,2,\ldots .$$

Whence

$$p_n = \frac{p_{n-2}}{1 + 2p_{n-2}} = \ldots = \frac{p_0}{1 + np_0}, \quad n = 0,1,2,\ldots .$$

This shows that p_n approaches 0 as n increases. Thus variation tends to be lost, in that the population will eventually comprise only *aa*-individuals. To determine how long it takes to reduce the fraction of *A* genes from p_0 to p, we have to solve for n the equation $p = p_0/(1 + np_0)$, giving $n = p^{-1} - p_0^{-1}$. For example, if $p_0 = \cdot 5$, it takes 8 generations to reduce the fraction to $\cdot 1$ and 998 generations to reduce it to $\cdot 001$. Although the *A*-gene must disappear, it may take a long time.

Case (b) $w_{AA} = w_{aa} = 1$, $w_{Aa} = 2$.

Since individuals who have both genes *A* and *a* are fittest, it is likely that both genes will survive. We might expect the proportion p_n to stabilise to some value p^* as n increases. For example, if $p_0 = \cdot 9$, then $p_5 = \cdot 595$, $p_{10} = \cdot 513$, $p_{15} = \cdot 502$, $p_{20} = \cdot 5002$, suggesting that $p^* = \cdot 5$. In fact, starting from any p_0 between 0 and 1, p_n approaches $\cdot 5$ as n increases.

The importance of this example is that it shows that *both* genes are maintained in the population, although we no longer have the constant proportions given in the Hardy-Weinberg Law. Eventually the population will comprise *AA*, *Aa*, *aa* in proportions $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ respectively.

MUTATION

Another genetic factor we have to allow for is mutation. Suppose that it is possible for *A* genes to change into *a* genes. In any generation, an *A* gene has a chance v ($0 < v \leq 1$) of becoming an *a* gene. We will ignore the possibility of *a* changing into *A*. The recurrence relation for p_n is given by

$$p_n = p_{n-1}(1 - v), \quad n = 1,2,\ldots . \qquad (**)$$

This is derived by noting that the proportion p_n of *A* genes at time n is the proportion p_{n-1} at time $n - 1$ times the chance $1 - v$ that the *A*'s do not mutate. From (**) it is clear that

$p_n = (1 - v)^n p_0$, and hence that p_n approaches 0 as n increases. This confirms the intuitive result that the A gene must eventually be eliminated.

To assess how strong the mutation force is, we can compute the time n required to reduce the A gene fraction from p_0 to p. We solve for n the equation $p = (1 - v)^n p_0$, leading to $n = \log_e(p/p_0)/\log_e(1 - v)$. For illustration, we take $p_0 = \cdot 5$ and $v = 10^{-6}$ (a typical value of the mutation rate in man). The time to reduce the fraction to $\cdot 1$ is $1 \cdot 61 \times 10^6$ generations, and to $\cdot 001$ it is $6 \cdot 21 \times 10^6$ generations. Comparing these results with those obtained in case (a), we conclude that selection can be a much stronger force in eliminating (genetic) variation than mutation.

SMALL POPULATIONS

The previous models have assumed that the population we are studying is very large. We might then enquire what effects small population size have on the previous results. The mathematical models are now much more complicated to analyse, see [2], but the following results indicate what can happen.

We consider a population of N individuals, each classified as one of the three genotypes AA, Aa, aa. To keep things manageable, we keep track only of the number of A genes, and not how these are arranged in individuals. Our population can then contain any of $X = 0,1,\ldots,2N$ A genes, the remaining $2N - X$ being a genes.

Before describing our model, we digress briefly to review some properties of the *binomial* distribution. Suppose we perform a series of n coin-flips, in each of which we have probability p of throwing a head, and $q = (1 - p)$ of throwing a tail. Then the chance of tossing $X = k$ heads is given by

$$\text{Prob}(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,\ldots,n, \qquad (***)$$

where $\binom{n}{k}$ is the binomial coefficient, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. The random variable X is said to have a binomial distribution with parameters n and p. Let's return to our genetic problem.

The model for the number of A genes in the next generation is derived as follows. If in generation n the number of A genes is i, then the proportion is $p_n = i/2N$. To produce the genes in generation $(n + 1)$, we take a binomial sample of size $2N$ from a very large pool of genes which are A or a in proportions p_n, $1 - p_n$ respectively. Then the number X of A genes at time $n + 1$ has the binomial distribution (***), with $n = 2N$, $p = p_n$. It can be shown that the average fraction of A genes at time n is just the proportion p_0 in the first generation. We conclude that *on average*, the A gene frequency remains constant, in agreement with the Hardy-Weinberg Law. However, this situation is very deceptive. It turns out that in fact the proportions of A genes *must* ultimately be 0 or 1, so that the

population eventually comprises either all *A* genes or all *a* genes. Genetic variation is consequently lost!

The phenomenon of loss of variation due to finite population size is known as *random genetic drift*. Clearly, it is possible to sample a population of size $2N$ genes which contains no *A* genes, in the same way that our series of coin-flips could result in no heads.

CONCLUSIONS

In this article, we have presented some of the genetic models that are used to describe the genetic composition of populations. The precise nature of the interplay between selection, mutation and random drift is still under active research.

REFERENCES

[1] *Biological Science: the Web of Life*. Australian Academy of Science, Chapter 34.

[2] Ewens, W.J. (1969). *Population Genetics*, Methuen, London.

[3] Hardy, G.H. (1908). Mendelian proportions in a mixed population. *Science*, Vol.28, pp.49-50.

[4] Weinberg, W. (1908). Über den Nachweis der Vererbung beim Menschen. *J. Ver. Vaterl. Naturk. Württemb.*, Vol.64, pp.368-382.

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

PROBLEM SECTION

PROBLEM 4.5.1.

What is the minimum number of hits necessary to score exactly 100 on this rather unusual rifle target? (This is like the "knapsack" problem mentioned in J. Stillwell's article in *Function* 4, Part 3.)