GENE FREQUENCY, A PROBABILISTIC APPROACH.

By

Nageeb E. Goher
Dept. of Animal Production, College of Agriculture
Cairo University, Giza, Egypt.

The definition of gene frequency is really the definition of the probability that an event occur. In a population with one loci, for instance, having n_1 of A_1 allele and n_2 of A_2 allele; the probability of A_1 allele is equal to $\frac{n_1}{n_1 + n_2}$. The same definition was given by Li (1955). Crow and Kimura (1971), define the gene frequency or the proportion of A_1 alleles to the total number of alleles in the population in case of one locus and two alleles (A_1 and A_2) to be $\frac{2 N_{11} + n_{12}}{2 N}$, where N_{11} equals the number of individuals in the population with genotype A_1A_1, N_{12} is equal to the number of A_1A_2 individuals, and N equals the total number of individuals in the population.

As an easy introduction to the concept of gene frequency to biologists, population geneticists, and to statisticians, this work was carried out. Here one can consider a case of dipliod population under random mating at one locus shuffling in the Mendelian process two or more alleles. Also, the same can be extended to case of polypliod populations distributing two or more alleles at one loci only.

It is known that for a population to a specific loci, with two alleles and the proportion of one allele $A_1 = P_1$ and the second allele $A_2 = P_2$ then the frequency of all possible genotypes are as follows:

Genotypes	A_1A_1	A_1A_2	A_2A_1	A_2A_2
Frequency	P_1^2	P_1P_2	P_2P_1	P_2^2

Actually the above frequencies form a probability distribution with the sum of all terms equal unity. Also, each term of the distribution consists of two independent events. Now if one defines a random variable X equal to the number of A_1 allele possessed by each individual in the population and then calculate the expected value for X along with the variance of X; it can be easily shown the relationship between E (X) and the gene frequency estimates.

Case I:

Dipliod populations under random mating at one locus and two alleles A_1 and A_2. The probability distribution function will be:

Genotypes	A_1A_1	A_1A_2	A_2A_2
Value of X	2	1	0
Probability of X	P_1^2	$2P_1P_2$	P_2^2

The $E(X) = 2P_1^2 + 2P_1P_2 = 2P_1$.

The variance of $X = 4P_1^2 + 2P_1P_2 - 4P_1^2 = 2P_1P_2$.

The value of gene frequency A_1 is equal to $\frac{2P_1}{2}$ or P_1, with variance equal to $\frac{2P_1P_2}{4}$ or $\frac{P_1P_2}{2}$.

Case II:

Dipliod populations under random mating considering one locus and 3 alleles A_1, A_2, and A_3. The terms of the probability distribution are as follows:

Genotypes	A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3
		A_2A_1	A_3A_1		A_3A_2	
Value of X	2	1	1	0	0	0
Probability of X	P_1^2	$2P_1P_2$	$2P_1P_3$	P_2^2	$2P_2P_3$	P_3^2

The $E(X) = 2P_1^2 + 2P_1P_2 + 2P_1P_3 = 2P_1$.

The variance of $X = 2P_1 (P_2 + P_3)$. Here also, the gene frequency for A_1 allele equal P_1 and its variance equal $\frac{P_1 (P_2 + P_3)}{2}$.

Case III:

Consider one locus and two alleles A_1 and A_2 in a polypliod population (Tripliod). The p.d.f. will be:

Genotypes	$A_1A_1A_1$	$A_1A_1A_2$	$A_1A_2A_2$	$A_2A_2A_2$
		$A_1A_2A_1$	$A_2A_1A_2$	
		$A_2A_1A_1$	$A_2A_2A_1$	
Value of X	3	2	1	0
Probability of X	P_1^3	$3P_1^2P_2$	$3P_1P_2^2$	P_2^3

The $E(X) = 3P_1^3 + 6P_1^2P_2 + 3P_1P_2^2 = 3P_1$

The variance of X is equal to $3P_1P_2$. The frequency of A_1 allele is equal to $\frac{3P_1}{3}$ or P_1 and its variance = $\frac{3P_1P_2}{9} = \frac{P_1P_2}{3}$.

Case IV:

Consider one locus and multiple alleles (A_1, A_2 and A_3) in a triplíod population.

Table 1. Genotypes, value of X, and probability of X for a tripliod population with three alleles

Genotypes	Value of X	Probability of X
$A_1A_1A_1$	3	P_1^3
$A_1A_1A_2$	2	$P_1^2P_2$
$A_1A_1A_3$	2	$P_1^2P_3$
$A_1A_2A_1$	2	$P_1^2P_2$
$A_1A_2A_2$	1	$P_1P_2^2$
$A_1A_2A_3$	1	$P_1P_2P_3$
$A_1A_3A_1$	2	$P_1^2P_3$
$A_1A_3A_2$	1	$P_1P_2P_3$
$A_1A_3A_3$	1	$P_1P_3^2$
$A_2A_1A_1$	2	$P_1^2P_2$
$A_2A_1A_2$	1	$P_1P_2^2$
$A_2A_1A_3$	1	$P_1P_2P_3$
$A_2A_2A_1$	1	$P_1P_2^2$
$A_2A_2A_2$	0	P_3^2
$A_2A_2A_3$	0	$P_2^2P_3$
$A_2A_3A_1$	1	$P_1P_2P_3$
$A_2A_3A_2$	0	$P_2^2P_3$
$A_2A_3A_3$	0	$P_2P_3^2$
$A_3A_1A_1$	2	$P_1^2P_3$
$A_3A_1A_2$	1	$P_1P_2P_3$
$A_3A_1A_3$	1	$P_1P_3^2$

Table 1 (continued)

Genotypes	Value of X	Probability of X
$A_3A_2A_1$	1	$P_1P_2P_3$
$A_3A_2A_2$	0	$P_2^2P_3$
$A_3A_2A_3$	0	$P_2P_3^2$
$A_3A_3A_1$	1	$P_1P_3^2$
$A_3A_3A_2$	0	$P_2P_3^2$
$A_3A_3A_3$	0	P_3^3

From Table 1, the E (X) is equal to $3P_1$ and its variance $3P_1$ $(P_2 + P_3)$. The gene frequency for A_1 will be P_1 with a variance $\frac{P_1 (P_2 + P_3)}{3}$.

In conclusion, the general formula for the relationship between the frequency of an A_i allele and the random variable X_i, where X_i refers to the number of i the allele per individual = $\frac{E (X_i)}{m}$ which would indicate that $E (X_i) = m P_i$, where P_i is the frequency of the i^{th} allele and m is the number of chromosomes per individual. The variance of gene frequency will be many of the above cases, $\frac{P_i (1-P_i)}{m}$, where P_i and m are the same as above.

This method can be extended to more than 3 alleles or 3 chromosomes, it also can apply to all problems in the poultry, animals, and or human populations where single locus under random mating is involved.

References

Crow, J. F. and M. Kimura. 1970. An introduction to population genetic theory. Harper & Row.

Li, C. C. 1955. Population genetics. Chicago University Press.