

Chapter 2

The Population Laws

In the preceding chapter we have studied the Mendelian ratios among the offspring of a given type of family with respect to one autosomal or sex-linked locus with two alleles. There may be hundreds, thousands, or millions of families in a human population. Throughout this chapter it is assumed that the population is very large. We may raise questions such as: What is the percentage of the three genotypes (AA, Aa, aa) in the general population? What is the relative frequency of each of the six types of families? To answer questions of this nature we need, in addition to the Mendelian law, two other specifications, which refer to the population as a whole. One is the mating system among individuals, and the other is gene frequency. The autosomal genes are considered first; the sex-linked genes are discussed briefly toward the end of this chapter.

Random Mating

Consider an arbitrary initial population in which 50 per cent of persons are AA, 20 per cent are Aa, and 30 per cent are aa. This population may be designated by (.50, .20, .30) for short. It is assumed that these genotypic proportions hold for both males and females. If men and women are married completely at random *with respect to this particular locus*, then the relative frequency of the six types of families is given by the appropriate terms of the product $(.50 + .20 + .30)(.50 + .20 + .30)$. To facilitate the arithmetic, the product can be arranged into the form of a multiplication table, as shown at the top of page 16, but it is unnecessary for more experienced readers to do so.

Parents		AA	Aa	aa
		.50	.20	.30
AA	.50	.25	.10	.15
Aa	.20	.10	.04	.06
aa	.30	.15	.06	.09

As seen that among all the families in the population there are 15 per cent of the type AA × AA. This is also the probability that two random individuals will both be AA. When the relative frequencies of the mating types are determined by the (product) probability of independent events, we say that it is a system of *random mating*, or *panmixia*. It should be pointed out that the term *random mating* is a relative one. Here we mean that mating is random relative to the A, a locus under consideration such as the hemoglobin types, MN blood-types, and so on. At the same time, mating may not be random with respect to height, education, intelligence, economic status, etc. The assumption of random mating with respect to certain genetic factors is no contradiction to assortative mating with respect to certain social factors or some other biological characteristics.

Counting AA × aa and aa × AA as the same type, as we did in the preceding chapter, we see that its total frequency is 15 + 15 = 30 per cent. Similarly, we pool the frequencies of other reciprocal crosses. The six types of families and their corresponding frequencies in the population are listed in the left half of Table 2-1. To calculate the offspring of these families, we apply the Mendelian laws to each type of family separately. Thus, for the 4 per cent Aa × Aa families in the population, $\frac{1}{4}$ of their offspring will be AA, $\frac{1}{2}$ Aa, and $\frac{1}{4}$ aa. The offspring of the entire parental population are given in the right half of Table 2-1 in terms of relative frequencies. The next generation—the total offspring of all families—is seen to be (.36, .48, .16). The results obtained so far can be summarized by writing:

$$(.50, .20, .30) \rightarrow (.36, .48, .16)$$

The reader at this stage may not be able to see any necessary

TABLE 2-1. TYPES AND FREQUENCIES OF OFFSPRING FROM A POPULATION
 OFFSPRING FOR AN ARBITRARY POPULATION (.50, .20, .30)
 PRACTICING PANMIXIA

Parents		Offspring		
Type	Frequency	AA	Aa	aa
AA × AA	.25	.25	0	0
AA × Aa	.20	.10	.10	0
Aa × Aa	.04	.01	.02	.01
AA × aa	.30	0	.30	0
Aa × aa	.12	0	.06	.06
aa × aa	.09	0	0	.09
Total	1.00	.36	.48	.16

relationship between the genotypic proportions of these two generations. However, the relationship will become immediately obvious when we consider another parameter of the population.

Gene Frequency

Although there are three genotypes in the population (.50, .20, .30), there are only two kinds of genes, A and a; we may ask: What proportion of the *genes* in the population is A and what proportion is a? Now, the AA individuals have all A genes; the Aa individuals have 50 per cent A genes and 50 per cent a genes; the aa individuals have all a genes. Hence, the proportion of genes in the population that are A and a are, respectively,

$$p = .50 + \frac{1}{2}(.20) = .60 \quad q = \frac{1}{2}(.20) + .30 = .40$$

These are called *gene frequencies*. Thus, $p = .60$ is the frequency of gene A, and $q = .40$ the frequency of gene a. Generally, in a population (D, H, R) where $D + H + R = 1$, the gene frequencies are:

$$p = D + \frac{1}{2}H \quad q = \frac{1}{2}H + R$$

These are, of course, also the frequencies of the A gametes and a gametes produced by the genotypes of the population as a whole.

The random mating of individuals implies that any two gametes (from opposite sexes) are united at random. If there are $p = .60$ A gametes and $q = .40$ a gametes in the population, random union of these gametes will yield an offspring population with:

$$p^2 = .36 \text{ AA} \quad 2pq = .48 \text{ Aa} \quad q^2 = .16 \text{ aa}$$

as illustrated in Fig. 2-1. This explains the numerical results

		$p = .6$						$q = .4$			
		A A A A A A						a a a a			
$p = .6$	A	AA	AA	AA	AA	AA	AA	Aa	Aa	Aa	Aa
	A	AA	AA	AA	AA	AA	AA	Aa	Aa	Aa	Aa
	A	AA	AA	AA	AA	AA	AA	Aa	Aa	Aa	Aa
	A	AA	AA	AA	AA	AA	AA	Aa	Aa	Aa	Aa
	A	AA	AA	AA	AA	AA	AA	Aa	Aa	Aa	Aa
	A	AA	AA	AA	AA	AA	AA	Aa	Aa	Aa	Aa
$q = .4$	a	aA	aA	aA	aA	aA	aA	aa	aa	aa	aa
	a	aA	aA	aA	aA	aA	aA	aa	aa	aa	aa
	a	aA	aA	aA	aA	aA	aA	aa	aa	aa	aa
	a	aA	aA	aA	aA	aA	aA	aa	aa	aa	aa

FIG. 2-1. The gene frequencies and equilibrium proportions of the three genotypes in a random-mating population.

obtained in the preceding section. It follows that the three initial genotypic proportions (.50, .20, .30) have no particular significance in a random-mating population except for their role in determining the gene frequencies $p = .60$ and $q = .40$. Any other initial population with the same gene frequency will yield the same offspring population on random mating. The reader may spend a few minutes to satisfy himself by showing that the initial population (.30, .60, .10) on random mating will yield the same (.36, .48, .16) offspring population.

An important principle is established in this section, viz., *the random mating of individuals is equivalent to the random union of gametes*. This principle will be employed many times in later

chapters in dealing with more complicated situations in which a longhand enumeration of all the possible types of mating is very tedious.

Equilibrium Condition

The immediate next question is: What will be the composition of the offspring generation of the parental population

$$(p^2, 2pq, q^2) = (.36, .48, .16)$$

if random mating continues? Applying the principle just established above, we obtain the *gene* or *gamete frequencies*:

$$p = .36 + \frac{1}{2}(.48) = .60 \quad q = \frac{1}{2}(.48) + .16 = .40$$

Random union of these gametes will yield an offspring population (.36, .48, .16) again. This shows that the population will remain the same in the absence of other disturbing factors. Such a population is said to be *in equilibrium*. This law was established independently by Hardy and Weinberg in the same year (1908) and hence is known as the Hardy-Weinberg law of equilibrium (see Stern, 1943).

Although the equilibrium has been established by a short cut, it is much more satisfying to demonstrate it by the long method, as shown in Table 2-2. In addition, we shall need all the details in the table for later discussions. This table is constructed exactly the same way as before, only replacing the arbitrary population by the one in equilibrium. For instance, the frequency of parental mating types is obtained from the terms of $(p^2AA + 2pqAa + q^2aa)^2 = (.36AA + .48Aa + .16aa)^2$. We see that the parental population $(p^2, 2pq, q^2)$ yields an offspring population of exactly the same composition.

Whatever the initial genotypic proportion, the equilibrium condition $(p^2, 2pq, q^2)$ will be reached in one single generation of random mating and will remain so on continued random mating. It is due to this simple theorem that we may safely take a human population to be in equilibrium with respect to one autosomal locus in practical research.

Comparing Table 2-2 in this section with Table 1-1 on family laws, we see that the chief difference is that in the present case

there is the extra column for "frequency of mating." We may think of the frequency of mating as "weight" for the various types of families. Then, the total offspring of the six types of families, properly weighted, is (p^2 , $2pq$, q^2).

TABLE 2-2. MATINGS AND OFFSPRING IN A LARGE RANDOM-MATING POPULATION WITH GENE FREQUENCIES $p = .60$ AND $q = .40$

Type of mating (mother × father)	Frequency of mating	Offspring		
		AA	Aa	aa
AA × AA	$p^4 = .1296$	$p^4 = .1296$	0	0
AA × Aa	$2p^3q = .1728$	$p^3q = .0864$	$p^3q = .0864$	0
AA × aa	$p^2q^2 = .0576$	0	$p^2q^2 = .0576$	0
Aa × AA	$2p^3q = .1728$	$p^3q = .0864$	$p^3q = .0864$	0
Aa × Aa	$4p^2q^2 = .2304$	$p^2q^2 = .0576$	$2p^2q^2 = .1152$	$p^2q^2 = .0576$
Aa × aa	$2pq^2 = .0768$	0	$pq^2 = .0384$	$pq^2 = .0384$
aa × AA	$p^2q^2 = .0576$	0	$p^2q^2 = .0576$	0
aa × Aa	$2pq^2 = .0768$	0	$pq^2 = .0384$	$pq^2 = .0384$
aa × aa	$q^4 = .0256$	0	0	$q^4 = .0256$
<i>Reciprocals combined (Weinberg, 1908)</i>				
AA × AA	$p^4 = .1296$	$p^4 = .1296$	0	0
AA × Aa	$4p^3q = .3456$	$2p^3q = .1728$	$2p^3q = .1728$	0
Aa × Aa	$4p^2q^2 = .2304$	$p^2q^2 = .0576$	$2p^2q^2 = .1152$	$p^2q^2 = .0576$
AA × aa	$2p^2q^2 = .1152$	0	$2p^2q^2 = .1152$	0
Aa × aa	$4pq^2 = .1536$	0	$2pq^2 = .0768$	$2pq^2 = .0768$
aa × aa	$q^4 = .0256$	0	0	$q^4 = .0256$
Total	1.0000	$p^2 = .3600$	$2pq = .4800$	$q^2 = .1600$

Now let us consider the q^2 , or the 16 per cent aa individuals in the general population. From the last column of Table 2-2, we see that they come from three types of families. The per-

proportions from each type in as follows:

$$\text{From } Aa \times Aa: \frac{p^2q^2}{q^2} = \frac{.0576}{.1600} = 36\% = p^2$$

$$\text{From } Aa \times aa: \frac{2pq^2}{q^2} = \frac{.0768}{.1600} = 48\% = 2pq$$

$$\text{From } aa \times aa: \frac{q^4}{q^2} = \frac{.0256}{.1600} = 16\% = q^2$$

Note that these percentages are identical with the genotypic proportions in the general population. Of course, this result may be seen directly (without referring to the table) by considering the probabilities. An *aa* individual implies that each parent has at least one *a* gene. These parents may thus be designated as $Xa \times Xa$, where *X* is the undetermined gene. The probability that both *Xs* are *A* is p^2 ; that one is *A* and one is *a* is $2pq$; and that both are *a* is q^2 .

Observational Verification

Before proceeding with the subject, let us look at some observed data to show that the Hardy-Weinberg law does actually hold in human populations.

The MN blood typing in man is a very familiar case in which all the three genotypes are distinguishable. Numerous published data from all parts of the world could be cited as examples; the following one has been chosen because it involves large numbers from a single community, and, as will be seen later, it also gives the frequency of the types of matings. Matsunaga and Itoh (1958) reported the following findings from the mining town

Genotype	Observed number	Gene frequency (estimate)	Expected proportion	Expected number
MM	406	$p = .525$ $q = .475$.2756	408.44
MN	744		.4988	739.22
NN	332		.2256	334.34
Total...	$G = 1,482$	1.000	1.0000	1,482.00

Ashibetsu in Hokkaido, Japan: When all three genotypes are distinguishable, a sample of $G = 1,482$ persons from a random-mating population is equivalent to a sample of $2G = 2,964$ genes. Each of 406 MM individuals, traditionally referred to as M individuals, has two M genes, whereas each of the 744 MN individuals has one M gene. Hence, there are $2(406) + 744 = 1,556$ M genes among a total of 2,964. The frequency of the M gene is then $1,556/2,964 = .525 = p$, and that of the N gene is

$$1 - .525 = .475 = q$$

This is not the true (but unknown) gene frequency of the Ashibetsu population but the estimate of it based on the observed sample. In statistical literature the sample estimate is written as \hat{p} or some other symbol to be distinguished from the true value (parameter). We shall not go too much into the sampling theory, and it is not necessary to use a new symbol. It is always understood that anything calculated from the sample is a sample estimate and is subject to sampling error.

The expected proportions of the three genotypes are the terms of $(p + q)^2 = (.525 + .475)^2$ in accordance with the Hardy-Weinberg law. Multiplying these expected proportions by $G = 1,482$, we obtain the expected numbers. It is seen that these expected numbers agree very well with those observed. The "goodness of fit" may be tested by the usual chi-square method. The value of χ^2 in this sample is 0.06 with one degree of freedom, showing a very insignificant deviation (that is, good fit) between the observed and expected numbers.

The sampling estimate ($p = .525$ for this particular sample) varies from sample to sample. If another sample were taken from Ashibetsu, the p value would be somewhat different from .525. The sampling variance of p or q is:

$$V(p) = V(q) = \frac{pq}{2G} = \frac{.525 \times .475}{2(1,482)} = .000084$$

and the standard error of p or q is the square root of the variance:

$$s(p) = s(q) = \sqrt{.000084} = .0092$$

The 1,482 persons above actually consist of 741 couples. The six mating types and their numbers are as follows:

Mating type	Observed number	Expected proportion	Expected number
MM × MM	58	$p^4 = .0760$	56.3
MM × MN	202	$4p^3q = .2749$	203.7
MN × MN	190	$4p^2q^2 = .2487$	184.3
MM × NN	88	$2p^2q^2 = .1244$	92.2
MN × NN	162	$4pq^2 = .2251$	166.8
NN × NN	41	$q^4 = .0509$	37.7
Total.....	741	1.00	741.0

In calculating the expected proportions for the various types of mating, the value $p = .525$ is employed. The expected and observed number of the mating types agree very well, and there is little doubt that the population is panmictic with respect to the genetic factors M, N. This furnishes us an observational proof of the general validity of the assumption of random mating and equilibrium.

Sex-linked Genes

The original Hardy-Weinberg law for a large random-mating population is established for autosomal genes only, but it also applies with slight modification to sex-linked genes *when the population is in equilibrium state*. The equilibrium state occurs when among females the genotypic proportions are p^2 , $2pq$, q^2 as in the autosomal situation and among the males the proportions are simply p , q . This is demonstrated in Table 2-3, in which, for the sake of omitting the constant factor $\frac{1}{2}$, the daughters and sons are listed as two separate populations of offspring; in each population the proportions add up to unity.

Several observations may be made on the equilibrium condition. Each female has two X-chromosomes, and each male has only one. If there are equal numbers of both sexes in the population, then $\frac{2}{3}$ of the sex-linked genes are in the female population and $\frac{1}{3}$ in the male population. However, the gene frequencies

in the two sexes are equal, both being p and q . This is an essential condition for equilibrium.

The proportion of males that possesses a sex-linked trait is always greater than the proportion of females that is homozygous for the locus. For instance, if $q = \frac{1}{12} = 8.33$ per cent of the

TABLE 2-3. EQUILIBRIUM CONDITION FOR SEX-LINKED GENES IN A LARGE RANDOM-MATING POPULATION

Mother	Father	Frequency of mating	Daughters			Sons	
			AA	Aa	aa	A·	a·
AA	× A·	p^2	p^2	p^2	
Aa	× A·	$2p^2q$	p^2q	p^2q	...	p^2q	p^2q
aa	× A·	pq^2	...	pq^2	pq^2
AA	× a·	p^2q	...	p^2q	...	p^2q	
Aa	× a·	$2pq^2$...	pq^2	pq^2	pq^2	pq^2
aa	× a·	q^2	q^2	...	q^2
Total.....		1.00	p^2	$2pq$	q^2	p	q

males are color-blind, then the proportion of color-blind females is only $q^2 = \frac{1}{144} = 0.69$ per cent, less than $\frac{1}{10}$ of 1 per cent. This explains why there are so few color-blind females in comparison with males. The smaller the value of q , the greater the departure between female and male proportions. Hemophilia is much rarer than color blindness, so hemophilic females must be very rare indeed. In addition to the q and q^2 relationship, there might be other factors affecting the incidence of female hemophiliacs. So far, only a paucity of cases has been reported.

The Approach to Equilibrium

With respect to an autosomal locus, equilibrium condition is reached in one generation of random mating. Because of the asymmetry in the two sexes, equilibrium is in general not reached in one generation of random mating for a sex-linked locus, except in special cases, but is approached very rapidly in successive

generations of random mating. The details have been given elsewhere (11, 1055a), and only a brief mention of the main features of the converging process is given here.

Suppose that among female parents the gene frequency is $q_{xx} = .32$ and that among male parents $q_x = .56$. The genotype of the sons is entirely determined by that of mothers and has nothing to do with the fathers. Hence the gene frequency among the sons will be the same (i.e., $.32$) as that among the mothers. A daughter's genotype is determined half by the mother and half by the father; therefore the gene frequency among the daughters is the average of their parental gene frequencies, that is, $\frac{1}{2}(.32 + .56) = .44$. Continuing this process, we obtain the following results:

Generation <i>t</i>	Females q_{xx}	Males q_x	Difference $q_{xx} - q_x$	Pooled frequency $\bar{q} = (2q_{xx} + q_x)/3$
0	.32	.56	-.24	.40
1	.44	.32	+.12	.40
2	.38	.44	-.06	.40
3	.41	.38	+.03	.40
...	.40	.40	0	.40

We see that each male gene frequency is equal to the female gene frequency of the previous generation and each female gene frequency is the average of two parental gene frequencies. The absolute difference between the female and male gene frequencies is halved in each generation. The shuffling of genes from one sex to another does not change the average or pooled gene frequency of the two sexes, as shown in the last column. In the equilibrium state, the gene frequency is equal to \bar{q} in both sexes. In practice, it is safe to take a human population as in equilibrium, except in a newly intermixed race.

APPENDIX ON ALGEBRAIC SIMPLIFICATIONS

Those who have no occasion to use algebra tend to forget the elementary manipulations they learned many years ago. This appendix serves to review some of the relations encountered frequently in the rest of the book, so that they need not block the reader when he sees them. One identity that the reader must be familiar with is:

$$(a + b)(a - b) = a^2 - b^2$$

This we shall use very often.

First, consider the linear expressions. If p and q are two positive fractions and $p + q = 1$, we may draw a straight line of unit length and divide it into two parts, one of length p and the other of length q , so a linear expression can always be thought of as a length.

$$\begin{array}{r} p = 1 - q \qquad q = 1 - p \\ p + 2q = 1 + q = 2 - p \quad 3 + q = 4 - p, \\ p - q = 1 - 2q = -2(q - \frac{1}{2}) \end{array}$$

Next, consider expressions involving the second power. Many of the following expressions may be seen by drawing a *unit square* with each side divided into p and q , so the unit area is $1 = p^2 + 2pq + q^2$. Thus $(p + q) \times q = 1 \times q = q$ stands for the *area* of a rectangle of unit length and q width.

$$\begin{array}{l} pq = q(1 - q) = p(1 - p) = q - q^2 = p - p^2 \\ p^2 + pq = p(p + q) = p \quad \text{similarly, } q^2 + pq = q \\ p^2 + 2pq = 1 - q^2 = (1 - q)(1 + q) = p(p + 2q) \\ \qquad \qquad \qquad = p(2 - p) \qquad \qquad = p(1 + q) = p + pq \\ p^2 - q^2 = (p + q)(p - q) = p - q = 1 - 2q \end{array}$$

In certain types of problems we encounter expressions involving the third power:

$$\begin{array}{l} p^3 + p^2q = p^2(p + q) = p^2 \\ p^2q + pq^2 = pq(p + q) = pq \end{array}$$

In dealing with mating types, expressions involving the fourth power come up very often. Usually there is more than one way of simplifying such expressions.

$$\begin{aligned}
 p^2 + 2pq + q^2 &= q^2(p^2 + 2pq + q^2) = q^2 \\
 p^2 + 4pq + 4q^2 &= (p^2 + 2pq)^2 = (1 - q^2)^2 \\
 &= p^2(p^2 + 4pq + 4q^2) = p^2(p + 2q)^2 \\
 &= p^2(1 + q)^2 \\
 q^4 + pq^3 + \frac{1}{4}p^2q^2 &= \frac{1}{4}q^2(1 + q)^2
 \end{aligned}$$

ally, expressions may involve quantities other than p and the following, s is a positive fraction.

$$\begin{aligned}
 q(1 - s) - q(1 - sq) &= -spq \\
 p^2 + 2pq(1 - s) + q^2(1 - s)^2 &= (1 - sq)^2 \\
 pq(1 - s) + q^2(1 - s)^2 &= q(1 - s)(1 - sq)
 \end{aligned}$$

These few examples must suffice. After a certain amount of practice, the reader should be able to simplify similar expressions without difficulty. The meaning of an expression often gives a clue as to how it should be simplified.