Mendel's Law and the Hardy-Weinberg Theorem

Mendel's model for hereditary transmission assumes that an individual will carry a gene controlling a feature that might be"on" or "off", having two *alleles*, call them a and A. Depending on which combination the individual carries (AA, aa, or aA, which is equivalent to Aa) the feature will be present or not. In the language of genetics, we consider a *genotype* determined by the combination of two *alleles*, denoted a and A. There are three genotypes: aa,aA,AA. Assume we start with a population with proportions AA:aA:aa given by p:$2r$:q ($p + q + 2\ r = 1$), and mating is random, with each allele also chosen at random.

1. What is the composition of the next generation
2. What is the composition of the following generations?

The calculation is not hard, and yields a somewhat unexpected result, known as the Hardy-Weinberg theorem.

Genes are inherited from parents, each contributing one allele per gene. LetX_1 and X_2 be the genes of the parents. By our assumption,

$$P\left[X_k = AA\right] = p, P\left[X_k = aa\right] = q, P\left[X_k = aA\right] = 2r$$

Assume parents are randomly paired, hence X_1 and X_2 are independent. Let Y_i be the allele transmitted by parent i. The allele must be present in the parent's gene. Thus

$$P\left[Y_i = A\,|X_i = AA\right] = 1$$

$$P\left[Y_i = A\,|X_i = aA\right] = \frac{1}{2}$$

$$P\left[Y_i = A\,|X_i = aa\right] = 0$$

and similarly for $Y_i = a$. Using total probabilities, we have that

$$P\left[Y_i = A\right] = P\left[Y_i = A\,|X_i = AA\right] P\left[X_i = AA\right] + P\left[Y_i = A\,|X_i = Aa\right] P\left[X_i = Aa\right] + P\left[Y_i = A\,|X_i = aa\right] P\left[X_i =$$

$$= 1 \cdot p + \frac{1}{2} \cdot 2r + 0 \cdot q = p + r$$

and similarly,

$$P\left[Y_i = a\right] = 0 \cdot p + \frac{1}{2} \cdot 2r + 1 \cdot q = q + r$$

Again, the Y_i's are assumed to be independent of each other.

The offspring will have a gene Z whose probability distribution will be, by independence,

$$x = P[Z = AA] = P[Y_1 = A]\, P[Y_2 = A] = (p + r)^2$$

$$y = P[Z = aa] = (q + r)^2$$

$$2z = P[Z = aA] = P[Y_1 = A]\, P[Y_2 = a] + P[Y_1 = a]\, P[Y_2 = A] = 2(p + r)(q + r)$$

Obviously

$$1 - (p + r)^2 - (q + r)^2 = 1 - p^2 - q^2 - 2r^2 - 2pr - 2qr = 1 - (p + q + 2r)^2 + 2\left(r^2 + pq + pr + qr\right) =$$

$$= 2\left(r^2 + pq + pr + qr\right) = 2(p + r)(q + r)$$

since $p + q + 2r = 1$.

What about the following generations? The next will have probabilities

$$x' = (x + z)^2, y' = (y + z)^2, 2z' = 2(x + z)(y + z)$$

And now, we observe that

$$(x + z)\cdot 1 = (x + z)(x + y + 2z) = (x + z)[(x + z) + (y + z)] = (x + z)^2 + (x + z)(y + z)$$

Similarly

$$y + z = (y + z)^2 + (x + z)(y + z)$$

Consequently

$$(x' + z')(y' + z') = \left[(x + z)^2 + (x + z)(y + z)\right]\left[(y + z)^2 + (x + z)(y + z)\right] = (x + z)(y + z)$$

and similarly $(x' + z')^2 = (x + z)^2$, and $(y' + z')^2 = (y + z)^2$. Thus, successive generations will all have the same proportion of alleles from the second on.