MATH 448 — Lecture 15

§6.2: Counting offspring

In the mouse cross $Aa \times Aa$, the allele A gives agouti fur and is dominant, so each offspring is agouti with probability $\frac{3}{4}$. Consider a litter of four. The number of agouti offspring in the litter is a *random variable*.

The binomial distribution. Let X be the number of agouti offspring in a litter of four. The litters with exactly three agouti offspring are NAAA, ANAA, AANA, AAAN, where N means non-agouti. The offspring are independent, so each of these has probability $(\frac{3}{4})^3\frac{1}{4}$, and by the addition rule $P(X=3) = 4 \cdot \frac{27}{256} = \frac{108}{256} \approx .42$. In general, suppose there are n independent trials, each a *success* with probability p and a *failure* with probability $q \coloneqq 1-p$. Any particular sequence with k successes has probability $p^k q^{n-k}$. Let $\binom{n}{k} \coloneqq$ the number of sequences of n trials with exactly k successes. Then

$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k q^{n-k}. \tag{15.1}$$

To count $\binom{n}{k}$, choose the k trials that are successes one at a time: there are $n(n-1)\cdots(n-k+1) = \dfrac{n!}{(n-k)!}$ ordered choices, and each set of k trials is chosen in $k!$ orders, so

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}. \tag{15.2}$$

For $n=4$ and $p=\frac{3}{4}$, (15.1) gives the distribution

k	0	1	2	3	4
$P(X=k)$	$\frac{1}{256}$	$\frac{12}{256}$	$\frac{54}{256}$	$\frac{108}{256}$	$\frac{81}{256}$

The possible values of a random variable together with their probabilities form its *probability distribution*. The table above is the probability distribution of X.

Expected value. For a random variable X with possible values i, the *expected value* is the average of the values weighted by their probabilities,

$$E(X) \coloneqq \sum_i i\,P(X=i). \tag{15.3}$$

For the litter of four, $E(X) = \dfrac{0\cdot 1 + 1\cdot 12 + 2\cdot 54 + 3\cdot 108 + 4\cdot 81}{256} = 3$. For any random variables X and Y, $E(X+Y) = E(X) + E(Y)$. Write the number of successes in n trials as $X = X_1 + \cdots + X_n$, where $X_j = 1$ if trial j is a success and 0 if not. Then $E(X_j) = p$, so

$$E(X) = np. \tag{15.4}$$

Spread. The *variance* of X is $\mathrm{Var}(X) \coloneqq E\big[(X-E(X))^2\big]$ and the *standard deviation* is $\sqrt{\mathrm{Var}(X)}$. For a single trial, $\mathrm{Var}(X_j) = E(X_j^2) - p^2 = p - p^2 = pq$. The variances of independent random variables add, so for the binomial distribution

$$\mathrm{Var}(X) = npq. \tag{15.5}$$

Mendel's F_2 generation of tall and dwarf plants had $n = 1064$ plants, each tall with probability $p = \frac{3}{4}$. Then $E(X) = 798$ and the standard deviation is $\sqrt{1064 \cdot \frac{3}{16}} \approx 14.1$. He observed 787 tall plants, which is 11 below the expected value.

The χ^2 statistic. For observed counts O and expected counts E in each category, the χ^2 *statistic* is

$$\chi^2 \coloneqq \sum \frac{(O-E)^2}{E}. \tag{15.6}$$

For two categories, let X plants be tall and $n - X$ dwarf, with expected counts np and nq. The deviation for dwarf is $(n - X) - nq = -(X - np)$, so both terms of (15.6) have numerator $(X - np)^2$, and using $p + q = 1$,

$$\chi^2 = \frac{(X-np)^2}{np} + \frac{(X-np)^2}{nq} = (X-np)^2\,\frac{q+p}{npq} = \frac{(X-np)^2}{npq} = \Big(\frac{X-np}{\sqrt{npq}}\Big)^2.$$

By (15.5), $\sqrt{npq}$ is the standard deviation of X, so χ^2 is the square of the number of standard deviations by which X differs from $E(X)$.

For Mendel's counts, $n = 1064$, $np = 798$, $X - np = -11$ and $npq = 199.5$, so $\chi^2 = 121/199.5 \approx .61$. This agrees with $\dfrac{(787-798)^2}{798} + \dfrac{(277-266)^2}{266}$.

For large n, by the normal approximation to the binomial distribution, $(X - np)/\sqrt{npq}$ is approximately a standard normal random variable, whose absolute value exceeds 1.96 about 5% of the time. Hence $\chi^2 > 1.96^2 \approx 3.84$ occurs about 5% of the time when the model is correct.

Problem 15.1. A litter of four comes from $Aa \times Aa$. Find the probability that exactly two offspring are agouti, and the probability that at least one is not agouti.

Problem 15.2. Two carriers of the recessive Tay–Sachs allele have three children. Find the distribution of the number of affected children, and its expected value.

Problem 15.3. Each offspring of $Aa \times Aa$ is agouti with probability $\frac{3}{4}$, but $P(X = 3) \approx .42$ for a litter of four. Why is this not a contradiction?

Problem 15.4. Show that $\binom{n}{k} = \binom{n}{n-k}$ using (15.2), and again by counting sequences of trials.

Problem 15.5. Mendel counted 5474 round and 1850 wrinkled seeds in F_2. Find the expected counts and χ^2.

Answers. 15.1: $\frac{54}{256} = \frac{27}{128}$; $1 - \frac{81}{256} = \frac{175}{256}$. 15.2: $P(0), P(1), P(2), P(3) = \frac{27}{64}, \frac{27}{64}, \frac{9}{64}, \frac{1}{64}$; $E = \frac{3}{4}$. 15.3: $\frac{3}{4}$ is the probability for one offspring. Exactly three of four is one of five possible counts, and $3 = 4 \cdot \frac{3}{4}$ is the expected count, not a count that must occur. 15.5: expected 5493 and 1831, $\chi^2 = \dfrac{361}{7324 \cdot 3/16} \approx .26$.