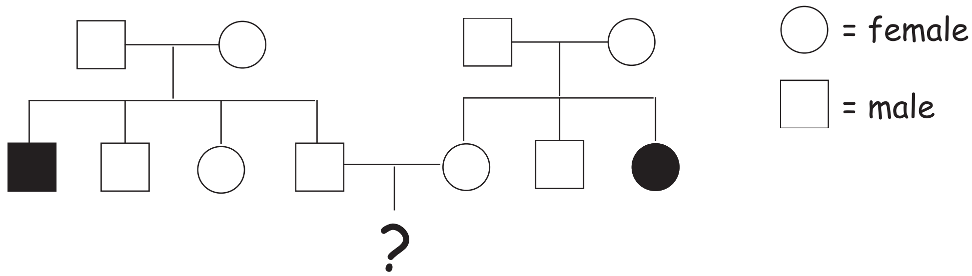
Lecture 5

Consider the following pedigree of an autosomal recessive trait.



p(affected child) = p(mother carrier and father carrier and affected child)

= 2/3 x 2/3 x 1/4 = 1/9

However, if they have a child that is affected we must reassess the probability that their next child will be affected.

p(both parents carriers) = 1. So, p(next child affected) = 1/4

This example shows how probability calculations are based on information. The probability changes not because the parents have changed but because our information about them has.

Now consider the case that the two parents have an unaffected child, with this new information we can recalculate the probability that the next child will be affected. An unaffected child does not establish definitively that both parents are **not** carriers, but it should be apparent that our estimate of the probability that both parents are carriers should be somewhat less than the original probability of 4/9. Correspondingly, the probability that the next child will be affected should be less than 1/9.

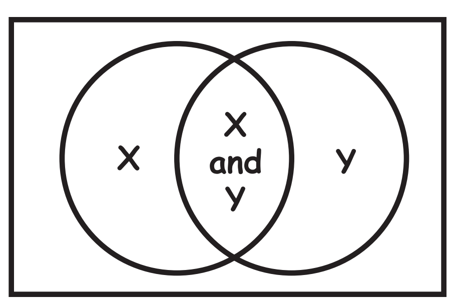
In order to calculate such probabilities exactly, we need to introduce a new concept known as **conditional probability**.

p(X|Y) = probability that event **X** will occur given that **Y** has occurred

$$\mathbf{p(X|Y)} = \frac{\mathbf{p(X \text{ and } Y)}}{\mathbf{p(Y)}}$$

Therefore:

p(X and Y) = p(X|Y) · p(Y)



Bayes Theorem

We can use this identity to perform a very useful type of probability calculation. In many simple probability calculations we can readily calculate the probability of a particular outcome (effect) given a particular circumstance (cause). Bayes' theorem allows us to calculate the inverse probability of a particular cause based on an observed effect. This method depends on knowing the probabilities of the measured event occurring given each of the possible causes and on knowing the a priori probabilities of obtaining the conditions for each of the possible causes.

Stated in terms of conditional probabilities, Bayes' theorem allows us to express **p(X|Y)** in terms of **p(Y|X)**, where **X** represents a particular circumstance (cause) and **Y** represents an observed effect.

$$p(X|Y) = \frac{p(X \text{ and } Y)}{p(Y)} = \frac{p(Y|X) \cdot p(X)}{p(Y)}$$

Consider a simple case in which there are only two possible circumstances/hypotheses (ie both parents are carriers or both parents are not carriers). We will express the complement of **X** (not **X**) as $\overline{X}$. Thus:

$$p(Y) = [p(Y|X) \cdot p(X)] + [p(Y|\overline{X}) \cdot p(\overline{X})]$$

$$p(X|Y) = \frac{p(Y|X) \cdot p(X)}{[p(Y|X) \cdot p(X)] + [p(Y|\overline{X}) \cdot p(\overline{X})]}$$

To apply this formula to the pedigree problem we will define **X** = both parents carriers; and $\overline{X}$ = not both parents carriers; **Y** = first child unaffected.

Accordingly, $p(X) = 4/9$, $p(\overline{X}) = 5/9$, $p(Y|X) = 3/4$, $p(Y|\overline{X}) = 1$

$$p(X|Y) = \frac{3/4 \cdot 4/9}{(3/4 \cdot 4/9) + (1 \cdot 5/9)} = \frac{3/9}{3/9 + 5/9} = 3/8$$

The probability that the next child will be affected is 3/8 x 1/4 = 3/32 = 0.094, which is slightly less than the probability we calculated before the couple had an unaffected child 1/9 = 0.111

Bayes' theorem is a powerful tool to analyze a wide variety of interesting problems that extend well beyond the field of genetics. For example, we can use it to caculate the probability of a misdiagnosis of AIDS infection using a diagnostic test for HIV. Assume that the test is very good and has both a false positive and false negative rate = 0.005,

Expressing each of these terms as conditional probabilities we have:

p(Pos|Inf) = 0.995 (probability that an individual tests positive given they are infected)

p(Pos|NInf) = 0.005 (probability that an individual tests positive given they are not infected)

Given that the a priori probability that an individual in the US has AIDS is 0.001.

Probability that an individual in the US is infected **p(Inf)** = 0.001

Using Bayes' theorem:

$$\mathbf{p(Inf \mid Pos)} = \frac{\mathbf{p(Pos \mid Inf) \cdot p(Inf)}}{\mathbf{p(Pos \mid Inf) \cdot p(Inf) + p(Pos \mid NInf) \cdot p(NInf)}}$$

$$= \frac{0.995 \cdot 0.001}{(0.995 \cdot 0.001) + (0.005 \cdot 0.999)}$$

$$= 0.16$$

The remarkable conclusion is that although the AIDS test has a very low error rate (for individuals who test positive there is a 99.5% chance that they have the disease) when the test is used broadly, only a minority of the positive tests would actually be AIDS cases.