MAT 2379, Introduction to Biostatistics

Chapter 5. Independence

In Chapter 4, we discussed the concept of "mutually exclusive" events. These are events which can not occur in the same time, i.e. the event that they happen simultaneously is impossible.

In this section we will discuss another important concept which defines the relationship between two events. This is the concept of "independence".

Two events A and B are independent if the fact that one has occur does not influence the chance that the other one will occur. In some cases, one can decide immediately if two events are independent or not, on a common sense basis.

For instance, the probability of "coming into substantial wealth" is the same regardless of the "message in your fortune cookie". In this case the two events are independent since

$$P(\text{wealth}|\text{message}) = P(\text{wealth}).$$

In other cases, one can not decide on a common sense basis, whether two events are independent or not. Think of our example (see notes Chapter 3) with a group of 100 German men. Is there any relationship between the "blond hair" event and the "blue eyes" event? We can not answer this question directly. (Intuitively, the fact that we already know that a randomly selected person has blond hair increases his/her chances of having blue eyes.)

This is why we need a precise definition. We will say that two events A and B are **independent** if the conditional probability of the event B given A coincides with the probability of B:

$$P(B|A) = P(B)$$

Example 1. Suppose that in a group of 350 women over the age of 50, 214 have a lifestyle which does not include exercising on a regular basis and 75 already have symptoms of osteoporosis. Moreover, 49 women which do not exercise on a regular basis show symptoms of osteoporosis. Is the fact that a women has symptoms of osteoporosis independent of her lifestyle?

Let A be the event "women does not exercise" and B be the event "woman shows signs of osteoporosis". From this study we can deduce that $P(A) = 214/350$ and $P(A \cap B) = 49/350$. Therefore the probability that a woman which does not exercise of developing osteoporosis is

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{49/350}{214/350} = \frac{49}{214} = 0.229$$

which is higher that $P(B) = 75/350 = 0.214$. The fact that a women has symptoms of osteoporosis is not independent of her lifestyle. The fact that a woman does not exercise increases her chances of having osteoporosis.

Alternatively, two events A and B are independent if:

$$P(A \cap B) = P(A) \cdot P(B)$$

The advantage of this formulation is that it can generalized to more than two events. We say that events A, B, C are **independent** if A, B are independent, A, C are independent, B, C are independent, and

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

Example 2. Approximately 50% of the population is male, 68% drinks to some extent, and 38.5% drinks and is male. Is the person drinking status independent of gender? (in class.)

Example 3. Three independent diagnostic tests $T1, T2, T3$ are run on the same patient. The probabilities that these tests will give correct results are: 90%, 85%, respectively 80%.

a) What is the probability that all 3 tests will give correct results?

Let A be the event "test $T1$ gives a correct result", B be the event "test $T2$ gives a correct result" and C be the event "test $T3$ gives a correct result". The statement of the problem is clear: the events A, B and C are independent. Hence the desired probability is

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) = (0.90) \cdot (0.85) \cdot (0.80) = 0.612$$

b) What is the probability that at least one test will result in error?

The event "at least one test results in error" is the complement of the event D:" all tests give correct results". We calculated above $P(D) = 0.612$. We have

$$P(D') = 1 - P(D) = 1 - 0.612 = 0.388$$

c) What is the probability that exactly two tests will give a correct result?

For answering this question we need to draw the tree diagram associated with this experiment. Note that in this case the "branches" are not equally probable. (Diagram to be drawn in class.) On the diagram we will see that there are three possible events which contain exactly two tests with correct results. These are:

- event F:"$T1$ correct (A), $T2$ correct (B), $T3$ error (C')"
- event G:"$T1$ correct (A), $T2$ error (B'), $T3$ correct (C)"
- event H:"$T1$ error (A'), $T2$ correct (B), $T3$ correct (C)"

We have

$$P(F) = P(A \cap B \cap C') = P(A) \cdot P(B) \cdot P(C') = (0.90) \cdot (0.85) \cdot (0.20) = 0.153$$

$$P(G) = P(A \cap B' \cap C) = P(A) \cdot P(B') \cdot P(C) = (0.90) \cdot (0.15) \cdot (0.80) = 0.108$$

$$P(H) = P(A' \cap B \cap C) = P(A') \cdot P(B) \cdot P(C) = (0.10) \cdot (0.85) \cdot (0.80) = 0.068$$

Finally, the probability that **exactly** two tests will give a correct result is

$$P(F \cup G \cup H) = P(F) + P(G) + P(H) = 0.153 + 0.108 + 0.068 = 0.329$$