Independent Events

Two events are said to be *independent* if the outcome of one of them does not influence the other. For example, in sporting events, the outcomes of different games are usually considered independent even though that may not be true in a completely strict and literal sense.

The *multiplication rule for independent events* says that if $A$ and $B$ are independent,

$$P(A \text{ and } B) = P(A) \times P(B).$$
Example: A football pundit states that the probability that UNC will beat NC State is 0.4, while the probability that UNC will beat Duke is 0.8. What is the probability that

1. UNC wins both games?
2. UNC wins at least one game?
3. UNC loses both games?
Solution:

1. If $A$ is the event “UNC beats State” and $B$ is the event “UNC beats Duke”, and if we assume these are independent events, then the probability of $A$ and $B$ is $0.4 \times 0.8 = 0.32$.

2. Apply the Law of Addition:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.4 + 0.8 - 0.32 = 0.88.$$  

3. Apply the Law of Complementary Events: “UNC loses both games” is the complement of “UNC wins at least one game”, so its probability is $1 - 0.88 = 0.12$. 

Warning:

Don’t confuse the notions of “independent events” and “disjoint events”. Independence means that the outcome of one event does not influence the outcome of the other. Disjoint means that if one event occurs then the other cannot occur — the very opposite of independence!
Conditional Probabilities

Consider the example (page 239 of text, referring to the Wimbledon tennis tournament),

\[ A: \text{“Federer misses his first serve”} \]

\[ B: \text{“Federer misses his second serve”} \]

We are told that Federer misses his first serve 44\% of the time, and that of all the times he misses his first serve, he also misses his second serve 2\% of the time.

What, then, is the probability he has a double fault?

Logically, the answer is 2\% of 44\%, or \(0.02 \times 0.44\), which is about 0.009.
Now let us rephrase this in the language of conditional probability.

We are told that the event $A$ occurs 44% of the time, or in other words $P(A) = 0.44$.

We are also told that, given that $A$ has occurred, the event $B$ occurs 2% of the time. This is written in probability notation as

$$P(B \mid A) = 0.02.$$ 

The left hand side is read as *the probability of $B$ given $A$*. In this particular context, it would not make sense to talk about the probability of $B$ given $A^c$, though in other contexts, that would make sense (e.g. free throws in basketball).
The law of multiplication for conditional probabilities says

\[ P(A \text{ and } B) = P(A) \times P(B \mid A). \]

Note that if we just interchange the role of \( A \) and \( B \), we also get

\[ P(A \text{ and } B) = P(B) \times P(A \mid B). \]

Finally, if \( A \) and \( B \) are independent, we get \( P(A \mid B) = P(A) \) and \( P(B \mid A) = P(B) \) — that formalizes what is meant by saying that the outcome of one event does not influence the outcome of the other. But in that case, either of the last two formulas reduces to

\[ P(A \text{ and } B) = P(A) \times P(B) \]

as in our earlier formulation of the multiplication rule for independent events.
Here is another (more complicated) example.

Consider the game in which a player tosses a die twice, and we want to calculate the probability that the total of the two tosses is at least 10. Define the events

\(A\): The first throw is a 6.

\(B\): The first throw is a 5.

\(C\): The first throw is a 4.

\(D\): The total of the two throws is at least 10.

Note that if the first throw is less than 4, it’s impossible for the total to be 10 or higher. So

\[
P(D) = P(A \text{ and } D) + P(B \text{ and } D) + P(C \text{ and } D). \tag{1}
\]
Now $P(A) = \frac{1}{6}$. Given that $A$ has occurred, $D$ will occur if the second throw produces any of 4, 5 or 6, and the probability of one of those outcomes is $\frac{3}{6}$ or $\frac{1}{2}$. So we have

$$P(A) = \frac{1}{6}, \quad P(D \mid A) = \frac{1}{2}, \quad P(A \text{ and } D) = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}.$$ 

Similarly

$$P(B) = \frac{1}{6}, \quad P(D \mid B) = \frac{1}{3}, \quad P(B \text{ and } D) = \frac{1}{6} \times \frac{1}{3} = \frac{1}{18},$$

$$P(C) = \frac{1}{6}, \quad P(D \mid C) = \frac{1}{6}, \quad P(C \text{ and } D) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$ 

Therefore, (1) leads us to

$$P(D) = \frac{1}{12} + \frac{1}{18} + \frac{1}{36} = \frac{1}{6}$$
giving the same answer as in our earlier calculation.
A diagnostic test for a certain type of cancer has a 98% chance of giving the correct outcome. Among all patients who take this particular test, 0.3% have the cancer in question. You take this test and the result comes back positive. What is the probability that you actually have cancer?

(A) 98%
(B) 86%
(C) 41%
(D) 13%
Solution: Draw a tree diagram

- **NO CANCER**
  - Probability: 0.997
  - **NEGATIVE TEST**
    - Probability: 0.97706
    - **POSITIVE TEST**
      - Probability: 0.01994
- **CANCER**
  - Probability: 0.003
  - **POSITIVE TEST**
    - Probability: 0.00294
  - **NEGATIVE TEST**
    - Probability: 0.00006
If $A$ is the event “patient has cancer” and $B$ is the event “patient has positive test”, then

\[
P(A \text{ and } B) = 0.00294,
\]

\[
P(B) = 0.00294 + 0.01994 = 0.02288,
\]

\[
P(A \mid B) = \frac{0.00294}{0.02288} = 0.1285.
\]

So the answer is about 13%. To many people encountering this issue for the first time, it is very surprising that the chance of actually having the disease, given a positive test, is so low. But it illustrates the problem created by “false positives” in many diagnostic settings.