Conditional Probability

1. Conditional probability basics
    a. With **conditional probability**, we are interested in <u>the probability that some event A happens given knowledge that some other event B happens</u>. For example, what is the probability that a random person is a student at IU given that they are a resident of Bloomington?
    b. We use a special notation for conditional probabilities: **P(A | B)** denotes the probability of event A given that event B occurs. So, in our example, if A is the event that a person is an IU student and B is the event that the person is from IU, P(A | B) denotes the probability that we are interested in.
    c. The conditional probability P(A | B) is given by the following formula:
        $\mathbf{P(A \mid B) = P(A \cap B) / P(B)}$
2. Why tree diagrams work
    a. We've learned to use tree diagrams to calculate the probabilities of events, but we have yet to describe why they work. Specifically, why is it that the outcome of an event is equal to the product of the edge probabilities? The edge probabilities actually correspond to *conditional* probabilities, and the reason that we multiply them is due to the **product rule** for conditional probabilities.
3. Product rule for conditional probability
    a. The product rule for two events is:
        i. $\mathbf{P(A \cap B) = P(A \mid B) \cdot P(B)}$
    b. Note that this just restates the definition of conditional probability that we saw before in a slightly different form.
    c. The product rule also extends to *n* events:

$$\mathbf{P(A_1 \cap \ldots \cap A_N) = P(A_1) \cdot (A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdots \ P(A_n \mid A_1 \cap \ldots \cap A_{n-1})}$$

4. Law of Total Probability
    a. Another important identity when working with conditional probabilities in the **law of total probability**, which allows us to calculate the probability of some event A by considering the two cases of whether or not some other event E occurs. The law is given by:
    b. $\mathbf{P(A) = P(A \mid E) \cdot P(E) + P(A \mid \text{not } E) \cdot P(E)}$
5. Inverse or Bayes probabilities
    a. A conditional probability P(B | A) is called an **inverse** or **Bayes** (or sometimes ***a posteriori***) probability if event B precedes A in time. The important point to note is that <u>mathematically inverse probabilities are no different from ordinary probabilities</u>.
6. Independence
    a. Another important notion related to conditional probability is that of **independence**. Intuitively, two events A and B are independent if the probability that A occurs is no different whether or not B occurs, and vice versa. In other words, knowing that A occurs does not give us any information about whether or not B occurs, and vice versa.

b. An intuitive example of independence is an experiment where two fair coins are flipped on opposite sides of a room. Intuitively, the way that one coin lands should have no affect on the way the other coin lands, so the two events are independent.

c. Formally two event A and B are independent if and only if:

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

d. Another equation for independence, which works when $P(B) \neq 0$, is given by

i. **$P(A \mid B) = P(A)$**

e. This equation is simply a way of rewriting the earlier equation, but makes the notion of independence a bit clearer. The equation states that the conditional probability of A given that B occurs is just the probability that A occurs. In other words, knowledge that B occurs does not change the likelihood of A.

7. Independence with more than two events

a. In order to extend the notion of independence to more than two events, we introduce the notion of **mutual independence**. Events E1, E2, …, En are **mutually independent** if and only if *for every subset* of the events, the probability of the intersection of those events is the product of the probabilities. In other words, all of the following equations must hold:

$P(E_i \cap E_j) = P(E_j) \cdot P(E_j)$ for all distinct i, j

$P(E_i \cap E_j \cap E_k) = P(E_j) \cdot P(E_j) \cdot P(E_k)$ for all distinct i, j, k

…

$P(E_i \cap \ldots \cap E_n) = P(E_1) \cdot \ldots \cdot P(E_n)$

b. As an example of mutual independence, tosses of 100 fair coins can be reasonably assumed to be mutually independent. However, as we can see from the equations, mutual independence is a very strong condition, and to check it requires a lot of calculations (since we must verify that every single equations holds!).

c. A weaker kind of independence between more than two events is **pairwise independence.** A set of events is pairwise independent if and only if every pair of events is independent. However, it is important to note that pairwise independence is a much weaker condition than mutual independence, and in general events that are pairwis independent are not necessarily mutually independent.