DEF: $P(A \mid B) \equiv$ the (conditional) Probability of A given B occurs

NOT'N: $\mid \equiv$ "given"

EX: The probability that event A occurs may change if we know event B has occurred. For example, if $A \equiv$ it will snow today, and if $B \equiv$ it is 90° outside, then knowing that B has occurred will make the probability of A almost zero. The probability of snow is higher if we do not know what the temperature is. Thus, $P(A \mid B) < P(A)$.

DEF: $P(A \mid B) = P(A) \equiv A$ is independent of $B \equiv$ the probability of A is unaffected by the occurrence of event B

EX: Consider two flips of a fair coin. $H \equiv$ Heads, and $T \equiv$ Tails.

$P(H$ 2nd flip $\mid H$ 1st flip$) = 1/2 = P(H$ 2nd flip$)$. That is, knowing the outcome of the first flip doesn't change the probability of the 2nd flip. So the two flips are independent.

NOTE: Conditional probabilities allow us to improve our estimates of probabilities by knowing more about the situation we are in. In elections, for example, knowing how many people are members of each party helps us to improve the accuracy of predictions about who will win the election. In a court case, knowing more about the circumstances in which a crime was committed helps us judge the probability of innocence or guilt.

Conditional probabilities allow us to reduce our sample space to just outcomes in the event we are conditioning on. For $P(A \mid B)$, we are finding the probability of A when the sample space is restricted to B. In a Venn diagram of probabilities, we would look only inside the area of B, and we would expand the area of B (and everything in it) to be unity. Our total probabilities of events in B would be unity. $P(A \mid B)$ would now correspond to the size of A in B, i.e., $A \cap B$ scaled up by the same factor that makes the size of B unity.

TOOL: The following formulas define the mathematical behavior of conditional probabilities:

$$P(A \mid B) = \frac{P(A,B)}{P(B)} \equiv \frac{P(A \text{ and } B)}{P(B)} \equiv \frac{P(A \cap B)}{P(B)} \qquad \text{(always true)}$$

$$P(A \mid B) = P(A) \qquad \text{(if and only if } A \text{ and } B \text{ independent)}$$

$$P(A,B) = P(A)P(B) \qquad \text{(if and only if } A \text{ and } B \text{ independent)}$$

TOOL: Using the Law of Total Probability and the axiom that probabilities of all outcomes in the sample space sum to unity, we can derive additional equations for conditional probability.

$$P(A'|B) = 1 - P(A|B)$$

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

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \text{(Bayes' theorem)}$$

We might be tempted to think that $P(A|B) + P(A|B')$ equals one, but that is not true. Events B and B' are different universes. Think of one as planet earth, and think of the other as a galaxy far far away. (Our universe consists of just these two places.) The probability of A in the different world may be totally different.

What is the relationship between $P(A|B)$ and $P(A|B')$? We put our equations to work.

$$P(A|B') = \frac{P(A,B')}{P(B')} = \frac{P(A \cap B')}{1 - P(B)} = \frac{P(A) - P(A \cap B)}{1 - P(B)}$$

which yields an equation involving $P(A|B)$ but which requires that we know $P(A)$ and $P(B)$:

$$P(A|B') = \frac{P(A) - P(A,B)}{1 - P(B)} = \frac{P(A) - P(A|B)P(B)}{1 - P(B)}.$$

Another interesting form has $P(B|A)$ instead of $P(A|B)$ but still requires that we know $P(A)$ and $P(B)$:

$$P(A|B') = \frac{P(A) - P(A,B)}{1 - P(B)} = \frac{P(A) - P(B|A)P(A)}{1 - P(B)} = P(A)\frac{1 - P(B|A)}{1 - P(B)}$$

Note that this last equation says that if A and B are independent, which means $P(B|A) = P(B)$, then $P(A|B') = P(A)$, which means A and B' are independent.