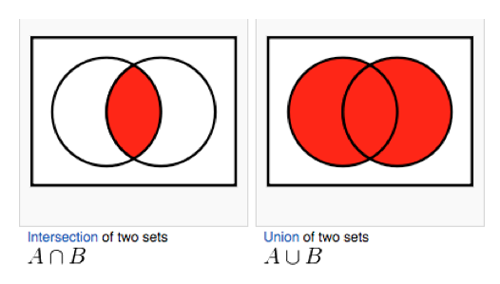
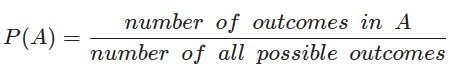
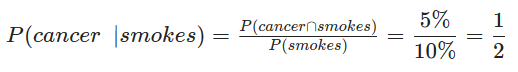
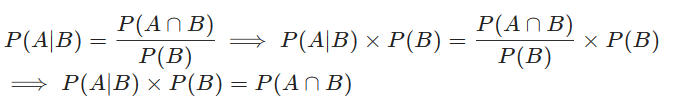
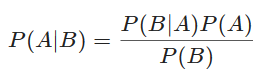
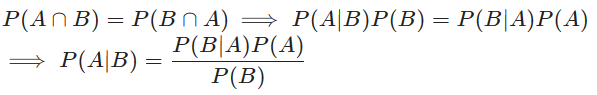
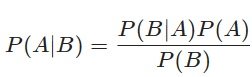
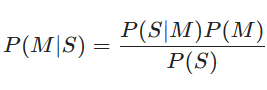
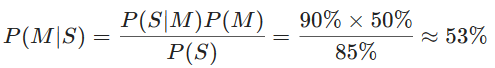
Unit 2-1 Axioms of Probability and Bayes’ Theorem

* Definitions
  + First, let's define a few terms that we'll refer to throughout the lesson.
    - **Probability**: The likelihood that an event will occur.
    - **Experiment**: A procedure that can be repeated an indefinite amount of times and has a well-defined set of outcomes.
    - **Event**: Any collection of outcomes of an experiment.
    - **Sample space**: The set of all possible outcomes of an experiment, denoted by *S*.
  + Examples
    - Suppose we flip a coin twice. What is the sample space? Define an event, E.
      * S = {{H,H}, {H,T}, {T, H}, {T, T}}
      * E = filp at least one heats = H, H, H, T, T, H
    - Suppose we roll a standard six-sided die once. What is the sample space? Define an event, E.
      * S = 1,2,3,4,5,6
      * E = roll an even number = 2,4,6
    - How would we denote these questions, mathematically?
      * Remember, an event is any collection of outcomes from an experiment. Sample space is the set of all possible outcomes of an experiment, denoted by S.
    - Suppose we flip a coin twice. What is the sample space? Define an event, E.
  + It's impossible to discuss probability without also addressing **sets** and **elements**. So what do these terms mean?
    - **Set**: A well-defined collection of distinct objects.
      * Example: The numbers 1, 2, and 3 are distinct objects but, when considered collectively {1,2,3}, they form a set.
    - **Element**: An object that is a member of a set.
      * Example: When considering Set A = {1,2,3}, the elements within A are the numbers 1, 2, and 3.
      * Example: When considering Set B = {1,2,(3,4)}, the elements of B are the numbers 1 and 2 and set {3,4}. In this case, there are three elements.
* Set Operations
  + Now that we're familiar with sets and their constituent elements, let's explore a couple of operations we can execute on them.
  + The union of two sets (A and B) is the set of elements that are in A or in B. We denote the union of A and B as A∪B.
  + The intersection of two sets (A and B) is the set of elements that are in both A and B. We denote the intersection of A and B as A∩B.
  + 
  + Set Operations Examples
    - Let A be the set of all even numbers between 2 and 9.
    - Let B be the set of all prime numbers between 2 and 9.
      * What is A∪B?
        + A∪B={2,4,6,8}∪{2,3,5,7}={2,3,4,5,6,7,8}
      * What is A∩B?
        + A∩B={2,4,6,8}∩{2,3,5,7}={2}
* Probability
  + Probability relies on sets.
  + We define events within a sample space and then evaluate the probability of that event occurring within that sample space.
  + Even when we don't realize it, we unconsciously follow this process when evaluating simple day to day probabilities, like the likelihood of finding a parking spot!
  + Now, let's dive in a bit deeper. A naive definition of probability would be to say that, for some event (A), the probability that A occurs is:
    - 
    - However, this isn't always correct — in fact, it's only true in cases where all outcomes are equally possible.
    - In other words, in order to define a sample space correctly, we have to make a series of assumptions. If we guess incorrectly, we compromise the accuracy of our predictions.
  + So any definition of probability is complicated by the need to formally define and unravel these assumptions, in order to avoid introducing error. Luckily, mathematicians and logicians have done some of this hard work for us by defining a few rules we can use to understand probability more accurately.
  + We call these rules the axioms of probability, a.k.a Kolmogorov's axioms.
  + Probability will always follow these axioms.
* The First Axiom of Probability
  + First: For any event: A, P(A) >=0
    - In other words, the probability of an event cannot be negative. OK, simple enough and makes sense, right? For instance, you'd never hear the probability of rain is "-12 percent."
* The Second Axiom of Probability
  + Second: For any sample space: S, P(S) = 1
    - That is, the probability of an outcome occurring in the sample space is 1. Think of this as a binary operation; if we assume probability exists on a scale from 0% to 100%, or 0 to 1.
    - Again, makes perfect sense, right? For instance, you'd never hear that there is a 101% chance of rain.
* The Third Axiom of Probability
  + Third: For mutually exclusive events:
    - 
  + This means that for a series of events that are mutually exclusive, the probability of a union of those events is the sum of the probabilities of the individual events.
  + Let's unpack that one a bit further.
    - Suppose we flip a coin.
    - Because we can't observe heads and tails simultaneously, these are mutually exclusive. Therefore, P(heads ∪ tails)=P(heads)+P(tails)
    - The probability of a union of those events is the sum (total) of the probability of each individual event.
* Marginal and Joint Probability
  + Moving on. The probabilities we've discussed so far have been marginal probabilities — probabilities of a single event.
  + For example, P(A) is the marginal probability of Event A.
  + However, we are often interested in the probability of two events occurring.
  + P(A∩B) is the joint probability of A and B. This can give us more information about how phenomena relate to one another.
  + For example, knowing the probability that an individual will develop cancer may be less relevant than knowing the probability that an individual will both smoke and develop cancer.
* Joint Probability
  + Of course, we can generalize joint probability to more than two events.
  + For example, if we want to estimate whether or not a given email is spam, we might be interested in the probability that the words "estranged," "prince," "bank," and "millions" appear in the same email.
  + We can use an intersection of sets to identify this. For example:
    - P(estranged ∩ bank ∩ prince ∩ millions)
* Conditional Probability
  + Slightly different from joint probabilities, conditional probabilities are used in cases where we know that an event has already occurred.
  + Suppose we're more interested in learning about the probability of cancer given that someone smokes. In other words, we know that smoking has already occurred.
    - We'd denote this as P(cancer∣smokes). That vertical bar — | — denotes smoking as a "given."
  + Similarly, if we're interested in whether or not an email is spam, given that our other trigger words were seen in an email, we'd find P(spam ∣ estranged ∩ bank ∩ prince ∩ millions).
  + The formula for conditional probability is an intuitive one.
    - If we want to understand the probability of *A given B*, then we'd look at where A and B happen together, out of all cases where we know B has already occurred.
    - Mathematically, we'd write this as:
      * P(A∣B)
    - In other words, we're restricting our sample space to only look at outcomes where B is true, then looking within that given space to see how frequently A also occurs.
  + Now suppose that 10 percent of individuals smoke and that 5 percent of individuals develop cancer and smoke.
    - Using this information, we want to find the probability that an individual who smokes will develop cancer.
    - 
    - How would we do this?
      * Assuming the previous information, we can state that the probability that someone who smokes will develop cancer is P(cancer∣smokes)=50%
* Conditional and Joint Probabilities
  + By using the formula for conditional probability, we can also easily find the formula for calculating joint probability.
    - 
    - The joint probability of *A and B, P(A* ∩ *B)*, is given by *P(A|B)* x *P(B)*
* Joint Probabilities
  + *P(A* ∩ *B) = P(A|B)* x *P(B)*
  + When finding the joint probabilities of multiple events, we find the probability of one event, then the probability of the next event given the first event, and continue on like this, using each event as a given to determine the probability the following event.
  + For example, suppose we want to find the joint probability of A, B, and C. Therefore, we have:
  + P(A∩B∩C)=P(C)×P(B∣C)×P(A∣B∩C)
  + This can generalize outward to an infinite number of events.
* Bayes' Theorem
  + **Bayes' theorem** is a powerful tool in statistics. We use it as the basis for an entire branch of statistics (called Bayesian statistics).
  + Bayes' theorem (also known as Bayes' rule) allows us to connect two related conditional probabilities:
    - P (A|B) *and* P(B|A)
  + We use the Bayes' theorem because it describes the probability of an event based on prior knowledge of conditions that might be related to the event. In other words, you're reasoning backwards to find the sequence of events that lead to a given outcome.
  + Bayes' statistics is different from frequentist statistics in that a Bayesian statistician would use **prior** or historic data as an input to their model, whereas a frequentist would use **new** information to make their conclusions
  + Before going forward, review the equation for the Bayes' theorem.
    - 
    - Suppose we have the probability that a smoker develops cancer. Using Bayes' theorem, we can reverse this and determine the probability that someone who has cancer (as a given) is also a smoker!
  + After all we've already learned about probability, the derivation of Bayes' theorem is surprisingly simple.
    - Let's recognize that P(A∩B) is the same as P(B∩A).
  + For example, let's assume that the probability that an individual smokes and develops cancer is equal to the probability that an individual develops cancer and smokes.
    - In this example, note that the order in which we write these two events is irrelevant.
    - Next, let's use our understanding of joint and conditional probabilities to derive Bayes' theorem:
    - 
    - The formal statement of Bayes' theorem is, for any two events (A and B):
      * 
  + Example
    - Suppose that you run a hotel with two restaurants: a Mexican restaurant and an Italian restaurant.
    - Based on recent customer reviews, 90 percent of Mexican restaurant patrons are satisfied and 80 percent of Italian restaurant patrons are satisfied.
    - For the purpose of this example, let's assume that 50 percent of our hotel guests eat at the Mexican restaurant while the other 50 percent of our guests eat at the Italian restaurant.
    - As you walk through the hotel lobby, you overhear a guest mention that her food was delicious. What is the probability that she ate at the Mexican restaurant?
    - How would we solve this? Remember, the best way to approach a complex task is to break down our steps into logical components. Spend some time thinking about this before advancing
      * Our first step is to write out the probability we seek to evaluate. We know that an individual was satisfied. We want to know how likely it was that he or she ate at the Mexican restaurant, specifically.
      * If we let S indicate "satisfied" and M indicate "dined at the Mexican restaurant," then we want to find P(M|S)
      * Using Bayes' theorem, we can find P(M∣S):
        + 
      * We can then fill in 50 percent for P(M) and 90 percent for P(S∣M).
        + These are pulled straight from the original facts — the probability that a randomly selected hotel guest eats at the Mexican restaurant is 50 percent and, of those who ate at the Mexican restaurant, 90 percent were satisfied.
      * The last piece to fill in is P(S), which represents the probability that a randomly selected hotel guest is satisfied with their meal. This includes both the Mexican restaurant and the Italian restaurant.
        + Finding P(S) is somewhat intuitive. Note that 90 percent of patrons of the Mexican restaurant are satisfied and 80 percent of patrons of the Italian restaurant are satisfied.
        + If you think that P(S) should be 85 percent, you're right!
      * The proper way to find P(S) is P(S) = P(S∩M) + P(S∩I) = P(S|M)P(M) + P(S|I)P(I) , where I indicates eating at the Italian restaurant.
      * However, because half of the hotel guests go to the Mexican restaurant and the other half go to the Italian restaurant, we can simply take the average of 80 percent and 90 percent.
      * This is called the **law of total probability**.
        + We can use the law of total probability in cases where a marginal probability is difficult to find. We will use this to calculate all of the joint probabilities that combine to make our marginal probability!
        + Moving forward with our problem, P(S)=85%.
      * Now, all that's left is to plug in our values and simplify things a little.
        + 
        + So what's our answer? [drumroll, please]
        + The probability that a satisfied hotel guest went to the Mexican restaurant is about 53 percent.
      * This example illustrates the benefits of conditional probabilities and, more generally, gathering evidence:
        + Without any evidence, we had a 50-percent chance of correctly guessing that someone went to the Mexican restaurant.
        + We know that patrons of the Mexican restaurant are more frequently satisfied than patrons of the Italian restaurant.
        + After knowing that the person was satisfied, we could precisely quantify how much likelier that person was to have gone to the Mexican restaurant.
      * Granted, the difference between 50 percent and 53 percent here seems small. But can you imagine adding evidence that, for example, allowed you to be correct 3 percent more frequently in the stock market?
      * Depending on the quantity and quality of the evidence you gather, you'll be able to make better *predictions based on probability*.
* Knowledge Check
  + Suppose that you run a lemonade stand and have two types of lemonade: sugar and non-sugar. Based on customer feedback, 95 percent of sugar customers are satisfied and 85 percent of non-sugar customers are satisfied. We also know that 75 percent of customers bought the sugar version and 25 percent bought the non-sugar version. You overhear a customer speaking fondly of their lemonade but forgot which one you sold them. What is the probability that the customer bought the sugar lemonade?
    - 77%
    - 