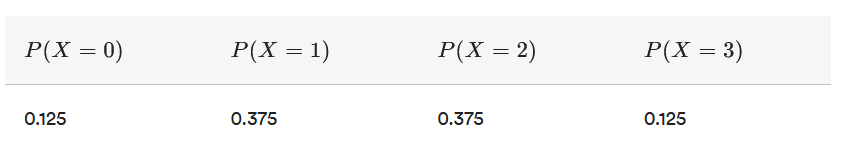
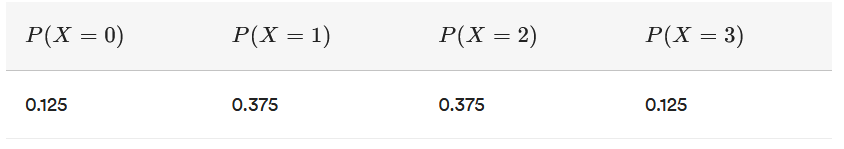
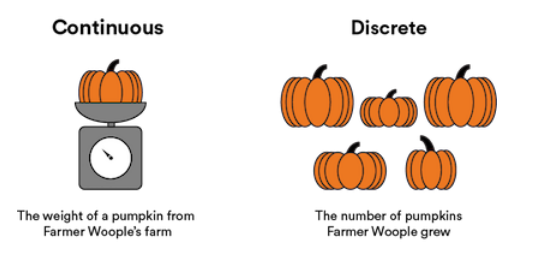
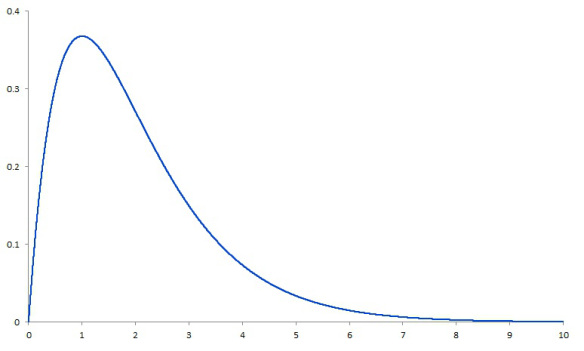
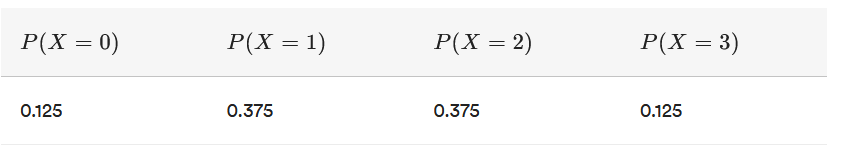
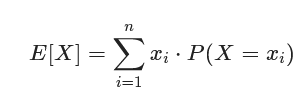
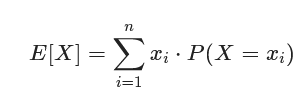
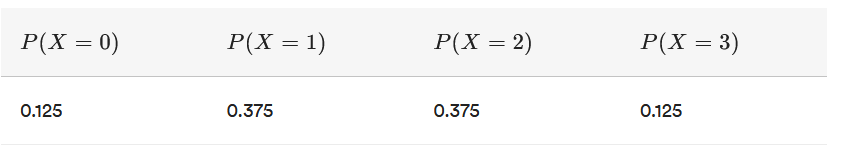
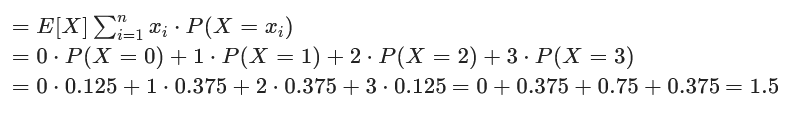
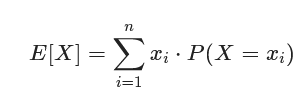
Unit 1-1: Random Variables and Expected Value

* Random Variable: the countable or numerical outcome of a chance event
* Introduction to Random Variables
  + Random variables are important for quantifying an outcome for a random process.
  + For example: Suppose you flip a coin three times and count the number of times the result is heads. Let XXX represent this count. We know that XXX can equal 0, 1, 2, or 3, because it is possible for us to flip zero heads, one head, two heads, or three heads. Again, XXX is a random variable here. It represents the outcome of your experiment (flipping three coins) with a number (how many heads you flipped).
  + Once a categorical outcome is quantified, we can derive probability and statistics from that process.
* Distribution of Random Variables
  + The distribution of a random variable describes all possible values of the random variable and how likely that variable is to take on each of those values.
  + We can see that the table here lists all possible values of our random variable XXX (0, 1, 2, 3) and how likely XXX is to take on each value (0.125, 0.375, 0.375, 0.125).
  + 
  + Back to our coin flip example — you flipped three coins and recorded the number of heads that came up.
    - Assuming this coin is fair and that each flip is independent of the others, we can derive the probability of each outcome:
    - 
    - We call this the distribution of the random variable.
  + If we know the distribution of a random variable, we can answer almost every question we could possibly have about XXX, like:
    - * If we were to flip three coins, what is the probability of observing at least two heads?
      * If we were to flip three coins, how likely is it that I observe exactly one head?
      * If we were to flip three coins, what is the probability of seeing either all heads or all tails?
      * If we were to flip three coins, what is the probability we will be served egg salad for lunch? (OK, so maybe not that one.)
  + Let's check out some real-world examples.
    - Suppose we record the number of free throws a basketball player shoots over a season and let YYY be the number of shots made.
      * If we were to analyze the data, what is the probability that a randomly selected player makes 80 percent or more of their attempted free throws?
      * If we were to analyze the data, what proportion of players make less than 60 percent of their shots and made at least seven of the first 10 they attempt?
      * See how a coach could use this information to better understand how well her players perform throughout a season?
    - Another scenario: Suppose we take the blood pressure of 10,000 adults. Let ZZZ be the blood pressure for each adult.
      * What proportion of adults have a systolic blood pressure above 140 and therefore suffer from hypertension?
      * What proportion of adults have a systolic blood pressure under 120 and are therefore in the "normal" range?
      * As you can imagine, the distribution of blood pressure can be very helpful for physicians and researchers to understand!
* Distribution of Random Variables
  + Nearly all random variables fit into one of two categories: continuous and discrete.
    - If you can list out all of the possible values of a random variable, we call that random variable **discrete**.
    - If you cannot list out all of the possible values of a random variable, we call that random variable **continuous**.
    - 
    - Note: There are a few fringe cases that don't fall into either category, but those are quite rare.
* Discrete Random Variables
  + Let's refer back to our earlier example of a random variable, XXX, where XXX is the number of heads flipped in a three-coin experiment.
    - In this case, XXX is discrete because you can write out all values of XXX (0, 1, 2, 3).
  + Now, consider our earlier example of a random variable, YYY, where YYY is the number of free throws made by a player.
    - In this case, YYY is discrete because you can write out all values of YYY (all of the possible shots made).
    - In a nutshell, if you can list out all values — even if it's a very large number with thousands of integers — then your random variable is discrete.
* Continuous Random Variables
  + Let's return to our earlier example of a random variable, ZZZ, where ZZZ is the systolic blood pressure of an individual.
    - In this case, ZZZ is continuous because you cannot write out all possible values of ZZZ.
    - For example, someone could have a systolic blood pressure of 120, or 120.1, or 120.01, or 120.001, and so on. You wouldn't be able to write out all possible values of their blood pressure.
  + Another example is a random variable that represents the distance between two objects. Because distance could be any value between 0 and infinity, this random variable would be considered continuous.
    - In the image below, as the x axis approaches 10, it appears that the y axis reaches 0. However, it will actually just become a smaller and smaller number on into infinity — 0.001, 0.0001, 0.00001, 0.000001, and so on.
    - 
* Summaries of Distributions
  + We generally represent a distribution with a table or histogram. These tell us a lot about the behavior of a random variable.
  + 
  + However, you might be interested in finding some ways to communicate information about this distribution to your boss without using these representations.
* Expected Value
  + We'll cover many different distribution summaries in the coming lessons, but for now we'll focus on expected value.
  + The expected value of a random variable is simply its average value.
  + We calculate expected value a bit differently for continuous or discrete distributions.
* Expected Value of Discrete Distributions
  + Take the random variable, XXX, where XXX is the number of heads observed when flipping three coins. We know from before that XXX can take on the values 0, 1, 2, and 3. The expected value of XXX — denoted E[X]E[X]E[X] — is 1.5.
    - It's not quite as easy as just averaging 0, 1, 2, and 3. In this case, it happens to work, but usually it won't. Consider if X=3X = 3X=3 occurred 90 percent of the time. In this case, do we think we'd want the expected value to be 1.5 or higher?
    - To calculate the expected value of some discrete random variable, XXX, we would use the following formula:
    - 
    - In this case, XXX is the random variable itself and XiX\_iXi​ are the different values of XXX. We take each value of XXX, multiply it by the probability of observing that particular value of XXX, and sum these products.
  + OK, recall the formula for...
    - 
    - ... Along with our table of values of XXX and their associated probabilities:
    - 
    - Now, let's calculate the expected value of XXX here:
    - 
* Definitions:
  + Calculus involves the act of finding integrals. To find the expected value of a continuous random variable, we'll need to integrate. In other words, we'll need to find the probability at each given point and solve for the cumulative sum of all those values.
* Expected Value of Continuous Distributions
  + Recall the formula for expected value applied to a discrete random variable:
  + 
  + In the case of a continuous random variable, we can't calculate  for each *i* — there's an uncountably infinite number of *i*s!
  + Rather than sum, we need to rely on calculus and integrate.
  + For a continuous random variable, XXX, the formula is , where *f(x)* is the probability density function for the random variable *X*.