1 Section 7.2: Frequency and Probability Distributions

When we have real life data for the outcomes of some experiment we can organize it into a **frequency table**. This is essentially a chart which gives each outcome and the number of times it occurs.

For example, the manager of a restaraunt wants to know how many times per week there is a wait for tables. After keeping track for 30 weeks the get the following table

Waits	weeks
3	1
4	5
5	6
6	8
7	8
10	2

This information all together is known as the frequency distribution. For all possible outcomes, we know the number of times it occurs. Now let's say we want to use this information to tell how successful the restaraunt is (or maybe understaffed). We also want to compare to another restaraunt, but they only had 20 weeks of data and they get

Waits	weeks
3	1
4	2
5	3
6	7
7	7
9	1

which is another frequency distribution. Can we say restaraunt 1 is busier because it's number's are bigger? That's not really fair because we only have 20 weeks of data. We can change this by instead working with a **relative frequency** table and distribution. For each table we divide each frequenc by the number of total outcomes, in this case 30 and 20. This information will tell us what proportion of the data falls in certain categores. For example, from the relative frequency distribution we can se the values for 6 and 7 are higher, and restaraunt 2 is probably busier.

This also gives us the emprical probability of each outcome. For example, the probability that restaraunt 1 has a line 10 times in a given week is 2/30. We can use our old tricks to find probabilities for more complicated events by adding.

We now turn to more thoretical experiments. In any experiment we have the notion of a **probability distribution**. This is the collection of all possible outcomes and their associated probabilities. For example, consider flipping a coin 3 times and recording the number of heads. In this case we can write down a table with each outcome and its probability. At that point we know everything about the experiment, and we can answer any probability question.

Example: We roll 3d4 and add the results together. Find the probability distribution for this experiment.

Roll	Probability
3	1/64
4	3/64
5	6/64
6	10/64
7	12/64
8	12/64
9	10/64
10	6/64
11	3/64
12	1/64

Now we can answer any question. P(roll odd), P(roll greater than 5), P(prime and single digit), P(roll 12—roll even)

We also define one of our key objects for the rest of the course. A **Random Variable** is a way to assign the outcome of an experiment to some number. For example, we roll 3d4 and denote the sum of the dice as X. Then X is a random variable. Note that this is often done in the most obvious way. We also use the notation $P(X = k)$ to denote the probability that a random variable X gives the outcome k. Then a random variable X has its own probability distribution

k	P(X=k)
x_1	$P(X = x_1) = p_1$
$\vdots$	$\vdots$
x_r	p_r

This notation is very convinent, and lets us write experiments in more familiar ways. For example, if Y, Z, W are all random vairalbes representing the outcome of a 4 sided dice, then we can write $X = Y + Z + W$ and we can say things like $P(Y > Z + 1)$ and manipulate algebraically.

Example: Suppose X is a random variable with $P(X = -1) = 1/3, P(X = 0) = 1/6$, and $P(X = 1) = 1/2$. Find $P(X^2 = 1)$.

Example: If we flip a fair coin n times and let X be the number of heads, describe the probability distribution for X. Now suppose the coin is weighted so we flip a head exactly p of the time. Do the same thing again.