To understand why we need probability in statistics goes back to the importance of randomly assigning treatments or randomly selecting samples from a population. RANDOM in this class means that an exact outcome is not predictable in advance, but a predictable long run pattern will emerge after many repetitions of an experiment or of a study (whose observations are selected at random). Examples might be: your commute home, how the stock market behaves, how people will respond to survey questions, how many are in class today.

PROBABILITY then is synonymous with the word CHANCE and it is a percentage or proportion of time some event of interest is expected to happen PROVIDED we have randomness and we can repeat the event (whatever it may be) many times under the exact same conditions (in other words – replication).

1. The Law of Averages AKA Law of Large Numbers

There is something called the Law of Averages(or the Law of Large Numbers) which states that if you repeat a random experiment, such as tossing a coin or rolling a die, a **very large number of times**, (as if you were trying to construct a population) your outcomes, when averaged, should be equal to (or very close to) the theoretical average (a parameter).

Remember the quote "The roulette wheel has neither conscience nor memory". Think about his quote and then consider this situation:

If you have ever visited a casino in Las Vegas and watched people play roulette, when gamblers see a streak of "Reds" come up, some will start to bet money on "Black" because they think the law of averages means that "Black" has a better chance of coming up now because they have seen so many "Reds" show up. While it is true that in the LONG RUN the proportion of Blacks and Reds will even out, in the short run, anything is possible. So it is wrong to believe that the next few spins will "make up" for the imbalance in Blacks and Reds.

The roulette wheel has no memory (and no conscience) so it has no idea that the last say, 10 spins resulted in "Red". The chance is the same that it will land on "Red" on the 11^{th} spin. Eventually in the long run (over thousands upon thousands of spins) it will even out – but remember, we live in the short run.

Basically then, if we think of each spin (or flip or attempt or draw or whatever unit you are studying) is a "trial". The larger the number of trials, the more likely it is that the overall fraction of "successes" will be close to the theoretical probability of a success in a single trial. Also with more trials: You are likely to miss the expected number of outcomes by a larger amount as measured by raw numbers, but you are likely to miss by a smaller amount in terms of percentages.

Example: A family has 7 children, all girls, and they really want a boy. Some people think they should try again because the "law of averages is going to kick in and the chance that the will have boy is really high" while other people think this particular family has the same chance (50%) of having an 8^{th} girl. Who is right? Why?

Many random processes (e.g. drawing a sample from a population, rolling a die, having kids, your commute, traffic conditions) are subject to this law.

2. Probability

A probability is

- A number between 0 and 1 also written $0 \leq P(A) \leq 1$
- It is the chance of some event's occurrence (e.g. event A)
- It can be the long run proportion of an event's occurrence

- OR it can simply be a known proportion (a given)
- OR it can be an equally likely outcome (e.g. the chance of drawing a particular card from a deck of playing cards is 1/52, the chance of seeing a particular face on a die is 1/6)

3. The Basic Probability Rules

1. A probability is a number between 0 and 1 or the probability of some event is $0 \leq P(A) \leq 1$, probabilities are never negative and never greater than 1.

2. The "something has to happen rule" or more formally, a sample space S is the set of all possible outcomes. The sum of all possible outcomes must equal 1 or P(S) = 1

3. The "complement rule" is the probability that an event does not occur is 1 minus the probability that it does occur. $P(A^c) = 1 - P(A)$ or the probability that something does occur P(A) is $1- P(A^c)$ the probability that it does not occur

4. The addition rule: If two events have no outcomes in common, the probability that one or the other occurs is the sum of their individual probabilities. P(A or B) = P(A) + P(B) this also means that the two events A and B are MUTUALLY EXCLUSIVE ("disjoint")

5. The multiplication rule: If two events have common outcomes but do not influence each other, they are independent. For example, I roll two die, one die should not have an effect on the roll of the second die. In these situations you are trying to figure out the chances of two things happening together. The chance that two things will happen equals the chance that the first will happen **multiplied** by the chance that the second will happen. P(A and B) = P(A) * P(B)

4. Putting it all together: An example

Suppose you get a job as a "assisted sales-person" in a large electronics store (like Best Buy) and prospective buyers behave this way 60% of the time they don't want anything to do with you and 40% of the time they will talk to you. Three people are walking towards you, assume they don't know each other and have INDEPENDENTLY chosen to walk towards you. What are the probabilities of 0, 1, 2, or all 3 talking to you?

Outcome	All 3 talk	2 talk to you	1 talks to you	0 talk to you
Probability	$.4^3$	(.4*.4*.6)+ (.4*.6*.4)+ (.6*.4*.4)	(.6*.4*.6)+ (.6*.6*.4)+ (.4*.6*.6)	$.6^3$

Suppose that if you can get at least two people talking to you, you are guaranteed a sale. What is your chance of a sale given these probabilities?

5. Summary – Why do you need to know this?

Well, you might decide to become a professional gambler. Just kidding. Probability (or chance) is a tool, an important tool, for the understanding of statistics. Recall that ultimately we work with statistics and make inferences (definition: the act of generalizing from statistical sample data) about the value of population parameters (usually with calculated degrees of certainty). One can imagine the parameter as being some kind of "truth" or "theoretical value" or a "long run value" and individual samples are a "short run" phenomenon and can be quite different from the "long run value", but in a somewhat predictable way (that's for later).