INTRODUCTION TO PROBABILITY

Unit 3A: Probability

UF | UNIVERSITY of FLORIDA

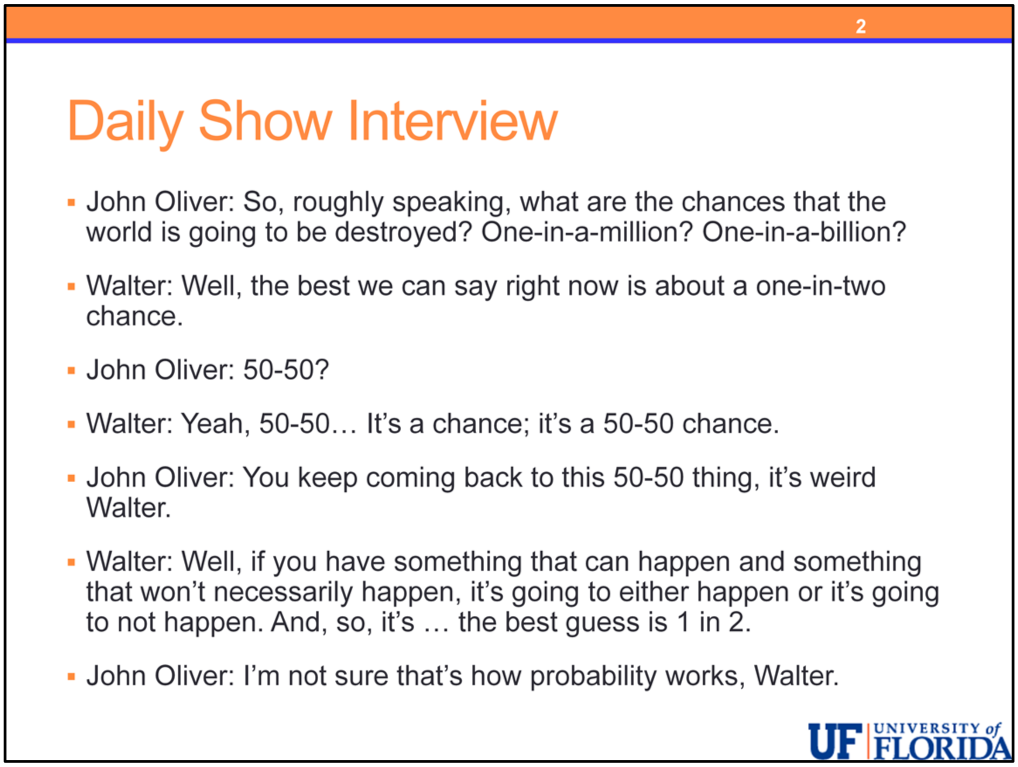
The course materials begin with two examples:

- The Let's Make a Deal problem and
- The Birthday Problem

Don't spend too much time on these examples or worry over their solutions. They are meant to get you thinking and to illustrate that what you THINK about the solution to a probability problem may not be accurate. Probability is not always intuitive.

In our course, we are not interested in studying complex probability rules or in solving problems like this, instead we are interested in estimating probabilities from data and understanding the probability concepts required for our study of inferential methods.

Probability can be thought of as the chance that something will happen or the likelihood that a particular event will occur.



Another point worth noting is one which was pointed out in an interview between John Oliver of the Daily Show with Jon Stewart and individuals related to the large hadron collider which slams particles together in the name of physics.

There were some news stories at the time that this device might create a black hole that would destroy the earth. Scientists say this is not going to happen but word gets out and spreads.

I suggest watching the clip to get the full effect but let's go through enough of John Oliver's conversation with Walter (a school teacher not a physicist).

John Oliver: So, roughly speaking, what are the chances that the world is going to be destroyed? One-in-a-million? One-in-a-billion?
Walter: Well, the best we can say right now is about a one-in-two chance.
John Oliver: 50-50?
Walter: Yeah, 50-50… It's a chance; it's a 50-50 chance.
John Oliver: You keep coming back to this 50-50 thing, it's weird Walter.
Walter: Well, if you have something that can happen and something that won't necessarily happen, it's going to either happen or it's going to not happen. And, so, it's … the best guess is 1 in 2.
John Oliver: I'm not sure that's how probability works, Walter.

And hopefully you believe that John Oliver is correct!!

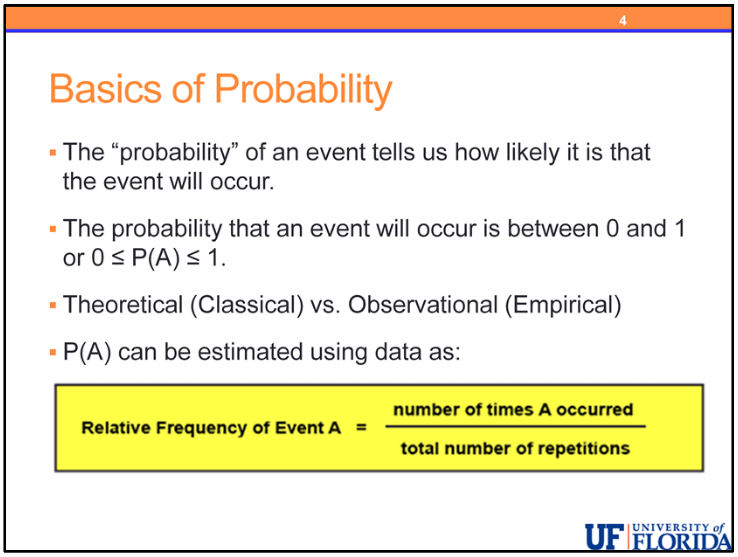
Some outcomes are equally likely – usually when they are completely random (flipping a fair coin, rolling a fair die, drawing cards from a shuffled deck, choosing people randomly from the population, etc.).

However, this is most definitely NOT the case for any pair of “opposite” events. Just because something either WILL or WILL NOT happen doesn’t mean the chance is 50-50.

Consider the chance you will develop a rare cancer. Is this 50/50?

How about the chance that you will get sunburned after being outside in July in a bathing suit with your skin unprotected for 1 hour? Is this 50/50?

Be very careful about the assumptions you make about scenarios involving probability. Assuming events are equally likely is a relatively common misunderstanding that is made about probability in certain situations.



The “probability” of an event tells us how likely it is that the event will occur. The probability that an event will occur is between 0 and 1 or $0 \leq P(A) \leq 1$. A probability of 0 implies the event will never happen and a probability of 1 implies the event is certain to happen. Usually events with probabilities of zero or one are not very interesting to discuss.

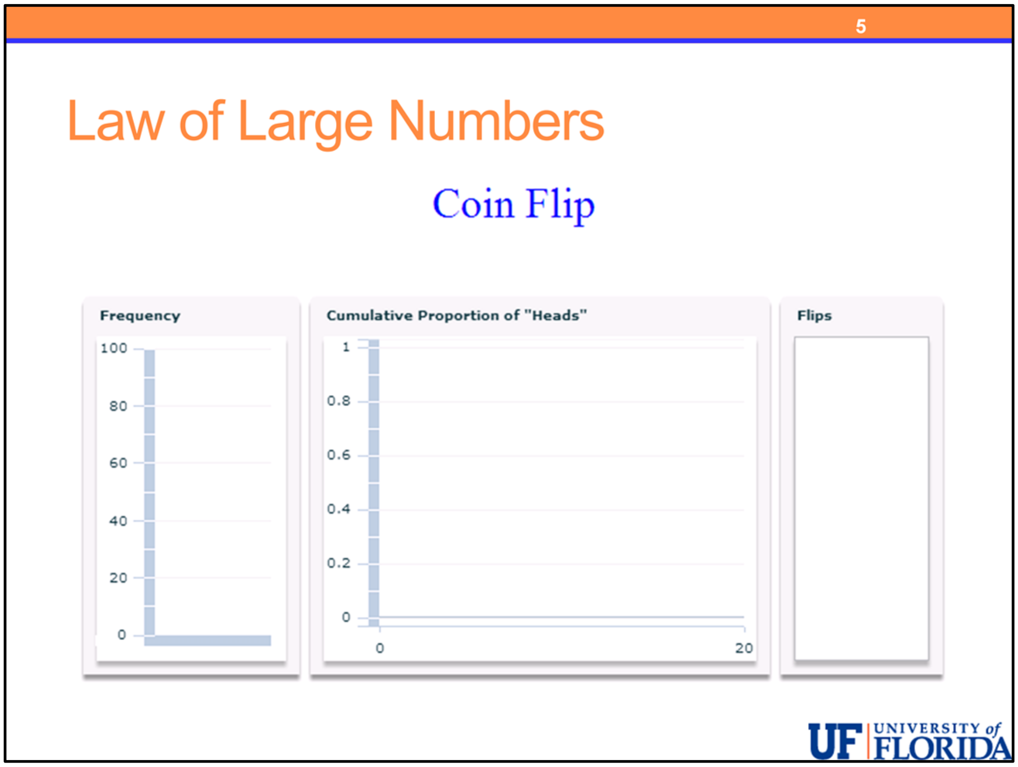
The notation P with something in parentheses after such as P(A) is used to represent the probability of A happening. We can include more descriptive notation such P(Male) or P(Have Diabetes) if we wish or we can use letters to represent our events of interest as a short-hand notation.

There are two ways to determine probabilities – one is by considering the situation theoretically such as games of chance. This is useful if it can be done, but we cannot possibly do this when we want to do something like estimate the prevalence of diabetes among the US population. In that case we must use data to estimate the desired probability empirically, based upon our observational data.

In this course we will occasionally approach probability from a theoretical perspective (more in Unit 3B than in Unit 3A) but for now we focus on determining probabilities based upon data. Some examples of theoretical probability are presented in the materials to help you understand the process of calculating probabilities but for the most part, you will be given data and asked to determine probabilities empirically.

To find probabilities from data, we simply need to count how many times our event occurred divided by the appropriate total. The appropriate total will be the entire sample size unless we are working with conditional probabilities which we will see are the same as the conditional percentage we covered in Unit 1.

This is a good time to point out that reviewing the materials for Case CC in Unit 1 would be very helpful as some of the questions are the same but with a different wording. Here we focus on probability language whereas before we called them percentages.



We will use the Interactive Applet on the Law of Large Numbers activity from the materials to illustrate.

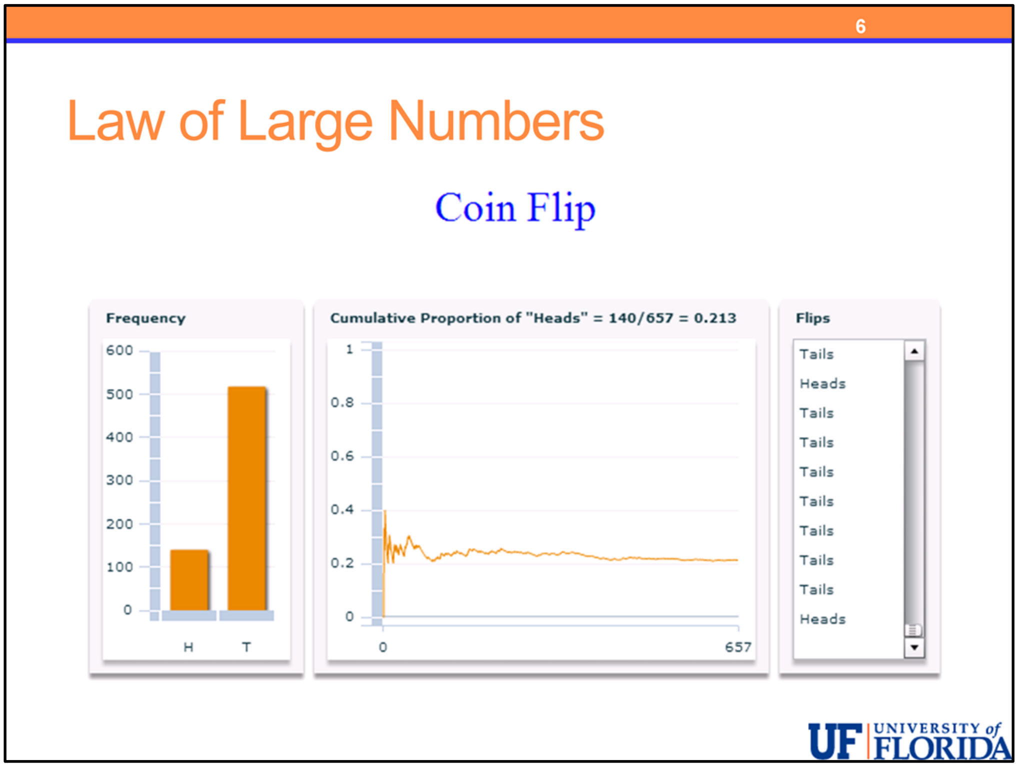
(Start Video Demo)

Here, we do not know if the coin is fair and can flip the coin as many times as we like to estimate the true proportion. I will begin by simulating one flip at a time, then I will increase it to 10 flips at a time and finally 100 flips at a time.

The law of large numbers says that the actual (or true) probability of an event (A) is estimated by the relative frequency with which the event occurs in a long series of trials.

In particular, if we continue to repeat the process, our estimate will, on average, get closer and closer to the true value (although it may go in the opposite direction for a while, eventually, it will continue toward the truth).

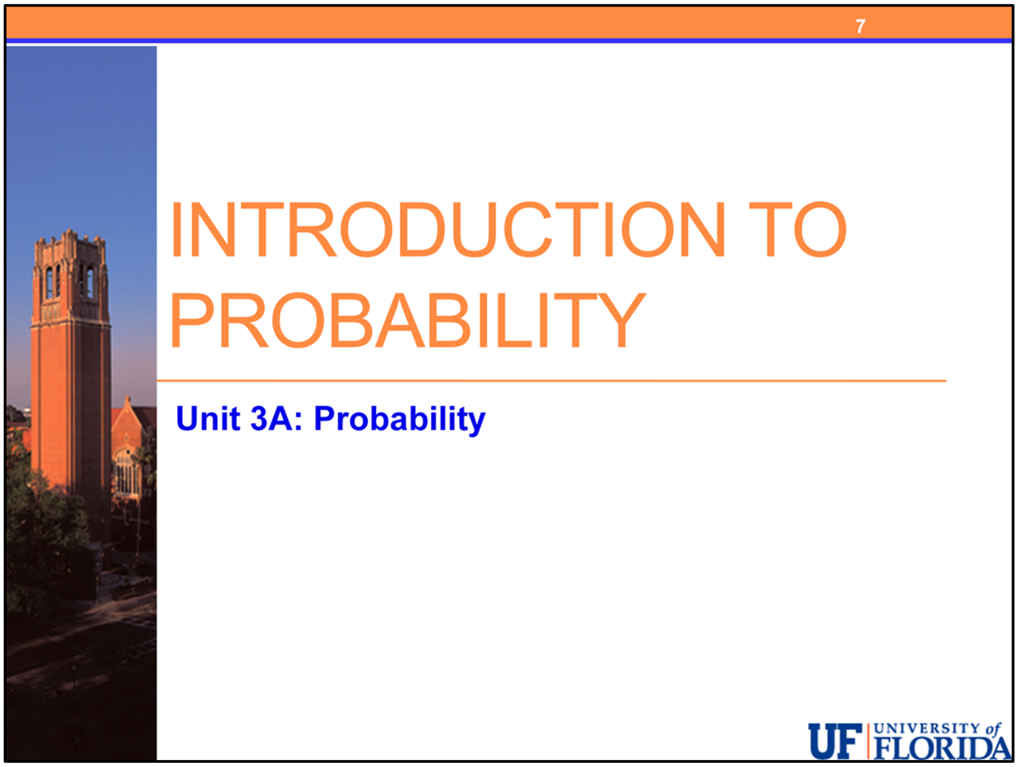
If this were not the case, there would be no reason to collect data to try to estimate probabilities. Because of this, the law of large numbers is a fundamental underlying rule used in almost all statistical methods in some way.



We can see after numerous flips that the true probability is clearly not 50%. Determining the exact probability is difficult but it seems to be around 20%. Later we will be able to construct a confidence interval and say something like:

Based upon 657 flips of the coin, the estimated proportion of heads is 21.3%. The 95% confidence interval suggests this value could be as small as 18.2% to as large as 24.4%.

From this simulation, you can probably convince yourself that as the number of flips increase, the variation in our estimate of the true proportion is decreasing. In the beginning, it varied wildly but early on it stabilized at a little above 0.2 and stayed very close, creeping toward 0.2 the larger the sample size became. We will study this process in Unit 3B when we discuss sampling distributions.



Review the examples in this section but remember not to worry about trying to solve our introductory examples on the let’s make a deal and birthday problems. These are only of interest to get you thinking and to illustrate that probability can be very un-intuitive!

In this section we simply want you to become familiar with the idea of relative frequencies to estimate probabilities from data and give you a feel for the law of large numbers.