In this segment we provide a high level introduction into the conceptual framework of classical statistics.

In order to get there, it is better to start from what we already know and then make a comparison.

We already know how to make inferences by just using the Bayes rule.

In this setting, we have an unknown quantity, theta, which we model as a random variable.

And so in particular, it's going to have a probability distribution.

And then we make some observations.

And those observations are modeled as random variables.

And typically we are given the conditional distribution of the observations given the unknown variable.

So these two distributions are the starting points, and then we do some calculations.

And we use the Bayes rule.

And we find the posterior distribution of theta given the observations.

And this tells us all that there is to know about the unknown quantity, theta, given the observations that we have made.

What is important in this framework is that theta is treated as a random variable.

And so it has a distribution of its own.

And that's our starting point.

These are our prior beliefs about theta before we obtain any observations.

However, one can think of situations where theta maybe cannot be modeled as a random variable.

Suppose that theta is at some universal, physical constant.

For example the mass of the electron.

Does it make sense to think of that quantity as random?

And how do we come up with a probability distribution for that quantity?

One can argue that in certain situations one should not think of unknown quantities as being random, but rather they are just unknown constants.

They are absolute constants.

It just happens that we do not know their value.

Or there may be other situations in which even though we may think that there is something random that determines theta, we are reluctant to postulate any prior distribution.

We do not want to impose any biases.

And that leads us to the classic statistical framework in which unknown quantities are treated as constants, not as random variables.

Pictorially the setting is as follows.

There's an unknown quantity that we wish to estimate.

And we make some observations, X. Those observations are random.

And they're drawn according to a probability distribution.

And that probability distribution depends, or rather is affected, by that unknown quantity.

So for example, for one value of theta, the distribution of the X's might be this one.

And for another value of theta, the distribution of the X's could be a different one.

And we're trying to guess what theta is.

Which in some ways is the question, do my data come from this distribution or do they come from that distribution?

In order to make a choice of theta, what we do is we take the data and we process them.

And after we process them, we come up with our estimate-- or rather estimator.

What is the estimator?

We take the data, and we calculate a function of the data.

That's what it means to process the data.

And that function is our theta hat.

Now this function, our data processing mechanism, is what we can call an estimator.

But quite often, or usually, we also use the same terminology to call theta hat itself an estimator.

Now notice that theta hat is a function of the random variable X. So theta hat is actually a random variable.

And that's why we denote it with an uppercase theta.

On the other hand, after you obtain some concrete data, little x, which are the realized values of the random variable capital X. Then we can apply your estimator to that particular input, and we compute a specific value-- call it theta hat lower case.

And that quantity we call an estimate.

So this is a useful distinction.

Always, with random variables, we want to distinguish between the random variable itself indicated by uppercase letters and the values of the random variable, which are indicated with lower case letters.

Similarly, the estimator is a random variable.

It's essentially a description of how we generate estimates.

Whereas the realized value, once we have some specific observations at hand-- that's what we call an estimate.

Now let me continue with a few comments.

The picture, or the setting, that I have here suggests that X is just one variable and theta is one variable.

But we can have the same framework, even if X and theta are multi-dimensional.

For example, X might consist of several random variables.

And theta may be a parameter that consists of multiple components.

Now you may notice that this notation that we're using here is a little different from our traditional notation which was of this form.

In what ways is it different?

The main difference is that here, theta is not a random variable.

Theta is just a parameter.

So what we're dealing with, here, is just an ordinary-- not a conditional distribution.

It's an ordinary distribution that happens to involve, inside its description, some parameters theta.

Just to emphasize the point that these are not conditional probabilities, because theta is not a random variable, we use a semicolon instead of using a bar.

And since theta is not a random variable, we do not include it in the subscript down here when we talk about the classical setting.

The best way to think of the situation mathematically is that we're essentially dealing with multiple candidate models, as in this picture.

This could be one possible model of X. This could be another possible model of X. We have one such model for each possible value of theta.

And if, for example, I were to get data points that sit down here, then a reasonable way to make an inference could be to say, these data are extremely unlikely to have been generated according to this model.

This data are quite likely to have been generated by this model.

So I'm going to pick this particular model.

So even though we're not treating theta as a random variable, and we do not have the Bayes rule in our hands-- we can still see, at least from this trivial example, that there should be a reasonable way of making inferences.

And let me close with some comments on the different types of problems that we may encounter in classical statistics.

One class of problems are so-called hypothesis testing problems in which we're asked to choose between two candidate models.

So the unknown parameter, as in this example, can take one of two values.

So think of a machine that produces coins.

And coins are either fair or they have a very specific bias.

You want to flip the coin, maybe multiple times, and then decide whether you're dealing with a coin of this type or of that type.

There's another type of hypothesis testing problems which is a little more complicated, for example this one.

We have one hypothesis which says that my coin is fair, versus an alternative hypothesis in which my coin is unfair.

But notice that this hypothesis actually includes many possible scenarios.

There are many possible values of theta under which this hypothesis would be true.

We will not deal with problems of this kind in this segment, or in this lecture sequence.

Instead we will focus exclusively on estimation problems.

In estimation problems, the unknown parameter, theta, is either continuous or can take one of many, many values.

What we want to do is to design an estimator-- a way of processing the data-- that comes up with estimates that are good.

What does it mean that an estimate is good?

An estimate would be good if the resulting value of the estimation error-- that is the difference between the estimated value and the true value-- if that difference is small.

You want to keep that difference small in some sense.

Well one may need a criterion of what it means to be small.

And whether we want this in expectation, or with high probability, and so on.

This statement, to keep the estimation error small, can be interpreted in various ways.

And because of that reason, there's no single approach to the problem of designing a good estimator.

And this is something that happens more generally in classical statistics.

Typically problems do not admit a single best approach.

They do not admit unique answers.

Reasonable people can come up with different methodologies for approaching the same problem.

And there is a little bit of an element of an art involved here.

In general, one wants to come up with reasonable methods that will have good properties.

And we will see some examples of what this may mean.

But again, I'm emphasizing that there is no single best method.

So whereas the Bayes rule is a completely unambiguous way for making inferences, here, in the context of classical statistics, there will be some freedom as to what approaches one might take.