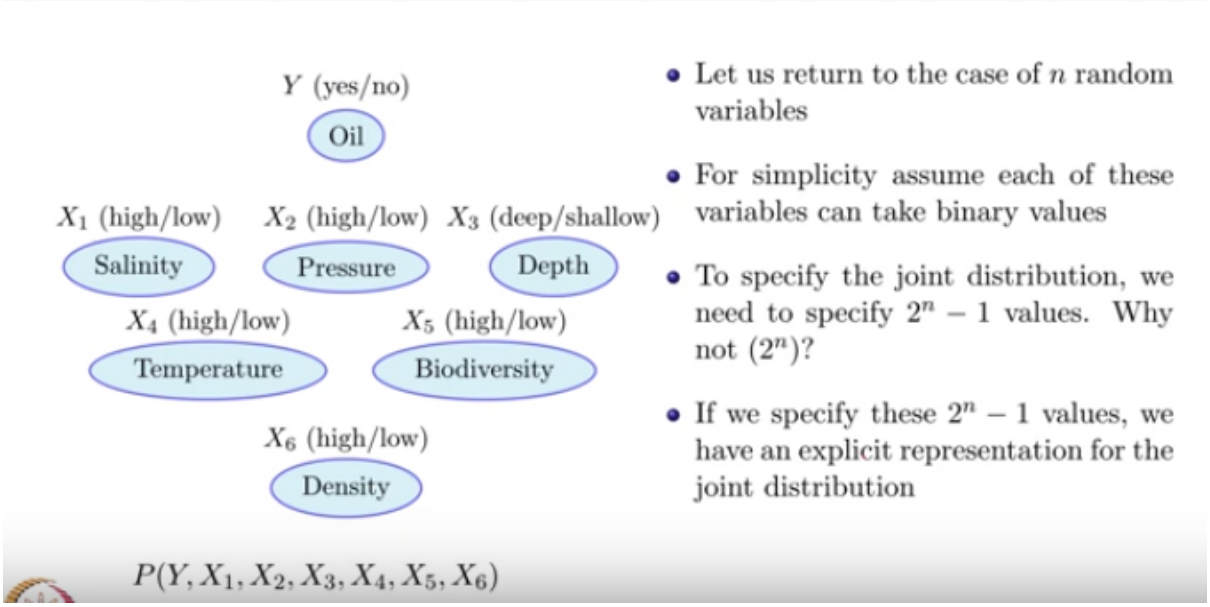
Lec 16.2

How do we represent a joint distribution

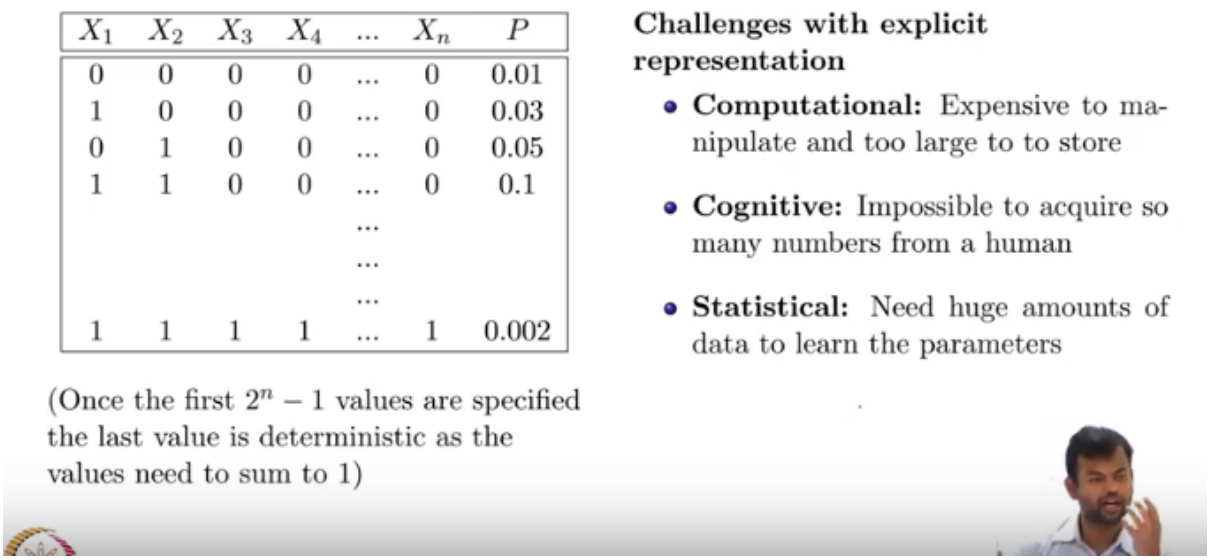
So now, Let's ask this Question. How do we represent a joint distribution? How do we represent a joint distribution? We have already done it in the past 20 minutes.

Refer Slide Time: (00:22)



Ok, so we will just look at that. So that is the question and it is good that some of you already realize it. It is just that, the question is asked in a way that, it is not very obvious, but we all of us have a regularly, routinely done this. Ok, so let us return to the case of 'End Variables 'and for simplicity let's assume that each of these variables can take binary values, right? 'Yes, No, high, low and so on, ok? And to specify the joint distribution and that is what the part of the recap was focusing on, if I have to give you the joint distribution, how many values do I have to give you? End variables, each of them is binary, if I have to give you a joint distribution, how many values do I have to give you? Two power N minus one, why minus one? It's obvious, so, if it is not obvious, it will come on the next slide. Right? It is always is there on the slide. So Y naught, two is to N, Y two is to N, it doesn't matter One value is not going to create a lot of difference. So if you specify these 2 is to N minus 1 value that is the representation of the joint distribution. That is how I choose to represent the joint distribution and I will give you a table, I will just put a value for every possible outcome in that joint distribution that means every possible combination of assignment to all the variables involved in the joint distribution and that is the representation. Is it fare? That is what I mean by representation, ok?

Refer Slide Time: (01:45)



Now, what is the problem with this explicit representation? Where I specify all the two is to N minus one variable, so I will tell you, there are three problems and you tell me which, what do I mean by these three words?

1. Computational

2. Cognitive

3. Statistic

Computational, what is the problem? just storing this table in your database or whatever is going to be a challenge, right? What is the Cognitive problem? Interpreting is one. Who is going to give you this table? So think of one situation, right? Again let us be practical in thing of the oil drilling case right? maybe there are some experts in the company who know, with various years of experience that how these things variables things interact with each other and just pull out numbers from their head right? Ok I think this looks like point zero three probability, this looks like point zero two and so on. And actually for small scale problems you can imagine that something like this is done, right. If you just have four or five variables, you could actually write it down, but as the value of N increases, what is going to happen? It is going to be impossible to acquire these tables or these values from humans, no human is going to able to give you that, if humans can't do it, who will do it? So now how will machines do it actually? they will learn from humans, they learn from data, so that brings me to the third part. What is the statistical problem? How many parameters do we have to learn? Two powers N. More the number of parameters, more the, more data, right? So to really learn these two days to end values, you will need more and more data, otherwise whatever you learn is not going to be statistically significant, right? well you have these two days to end possible entries in your table, atleast see each of these entries hundred times to be able to figure out, what their probabilities are right? It is just seeing them once itself is a challenge, how will you get so much of data? Ok, so these are the three problems right? You have, Cognitive, Computational, as well as Statistical problems, so this representation, although a valid representation, has its own set of problems. So, you can't really, even

though this representation is fine, you can’t really represent the joint distribution like this because it is going to be too many values, you can’t really do this representation.