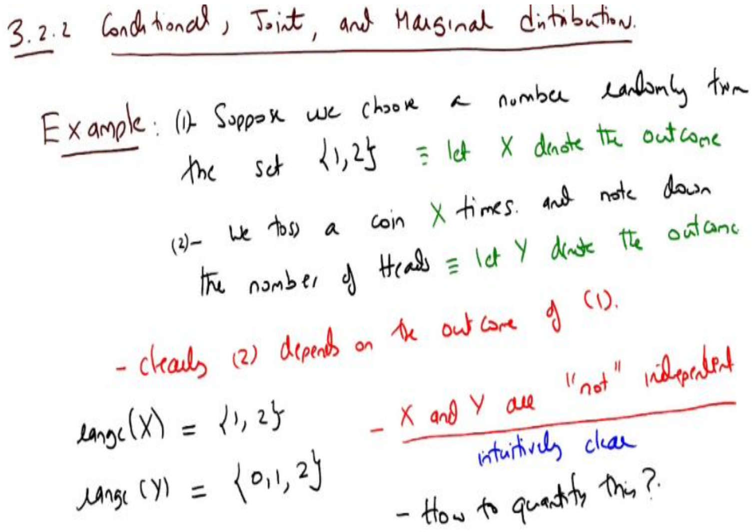
Introduction to Probability - With Examples using R
Professor Siva Athreya
Theoretical Statistics and Mathematics Division
Indian Statistical Institute, Bangalore
Lecture 16
Conditional, Joint and Marginal Distributions

So, we were doing so far independent random variables. And we did one computation about a bunch of random variables being independent, and how to use that an iid in fact they were all geometric P and they were able to use the fact that we use the use the kind of iidness to compute the probability that all in the begin equivalent j. This one standard computation, one can do in independence. So, now comes the next important concept, random variables are dependent. How does one understand? So, that whole concept is called conditional distributions.

(Refer Slide Time: 01:03)



So, section 3.2.2 in the book, these are called conditional distributions. Conditional, so there are 3 important concepts which I will do this week, conditional independence, and joint and marginal distributions. So, the idea is that, let us start with a simple example, start with simple example. So, we will start with an example. So, suppose we do a two-step experiment. So, we start with a simple example. I will not number it. So, suppose let us say, we chose a number randomly from the set say 1, 2 first step. So, let us say and let us do a computation let let us say let X denote the outcome.

And then what we will do is the second is step 1 of the experiments, let us say step 1 of the experiments. Step 2 of the experiment is you toss a coin X times, X times you toss a coin X times. And note down say the number of heads, note down the number of heads. So, here again, we say, let us say let us go here, and let us say let Y denote the outcome. So, how does one understand this? So clearly, we do, the experiment depends on each other, 2 depends on 1 very clearly. You choose a number between 1 and 2 and then, depending on the outcome you get, you toss a coin so many times.

So, clearly let us clearly the step one, may they did not appear clear I will use different colour so it is clear for you. So, clearly 1 and 2 or let us say or even this may not be that clear, clearly, but what is clear is 2 depends on the outcome of 1. So, then the idea is the following that so let us try and understand how to let us try and understand. So, first is range of X, range of X is just 1 and 2, range of Y, so that is a little tricky. So, X is 1, the range of X can be only 0 or 1. But when X is 2 range of X range of Y can all we have 0 1 or 2. And if you have 2 if you toss a coin twice then you can go and.

So, now clearly so the idea is that one important outcome of this experiment is that X and Y are not independent. That is something you have to, you have to think about. But so intuitively it is clear, because, clearly, depending on X values of Y change. So, in some sense, the X of X outcome of Y. Let us try and see, let us try and understand how to quantify this? How to quantify this? So here, one has to be, one has to just think a little bit and understand. So, X is easy to understand. X is just the distribution of X is easy to understand. So, let us see what is easy Let us see.

(Refer Slide Time: 06:51)

range $(Y) = \{0, 1, 2\}$

- How to quantify this?

(a) $\mathbb{P}(X=1) = \frac{1}{2} = \mathbb{P}(X=2)$ as $X \sim \text{Uniform}\{1,2\}$

(b) $\mathbb{P}(Y=0 \mid X=1) = 1-p$ ... (ie toss a coin once and outcome is a Tail)

$\mathbb{P}(Y=1 \mid X=1) = p$ ... (ie toss a coin once and outcome is a Head)

$\mathbb{P}(Y=0 \mid X=2) = (1-p)^2$ ... (ie toss a coin 2 times and no head appears)

$\mathbb{P}(Y=1 \mid X=2) = \binom{2}{1} p^1 (1-p)^{2-1}$

So, the chance that X is equal to 1 is the same, so what is all clear? Just write it down. So first come first is the following so the chance that X is equal to 1 is the same as half. And that is the same as chance at Y is equal 2, X is equal to 2. This X is uniform as X is uniform 1 comma 2. So, that is clear.

So, the distribution of X is easy to understand, now how do you understand the distribution of Y? So, what is Y? Y means, if you know, you get outcome X, you toss a coin, X times and you find the number of X. Let us say so then the idea is that so Y takes the value 0, 1 or 2. So what do you know what Y? So, first thing you know is, if Y is equal to 0 you need to understand what happened to X. So, let us say say X is equal to 1, so Y is equal to 0 given X is equal to 1, for me to toss a coin once and you want know, the chance that you get no heads, that is just the same as 1 minus P that is you toss a coin once and the outcome is a tail. That is one thing.

The other thing is that if X equals 1 Y can take the value 1 that is, we can get a head. So, Y equal to X equal to 1 is then same thing you can duplicate it, let us see how I can do this, copy this and bring it down very nice and so then also I erase this tail and write head, (())(09:00) that what we see here.

So, now what this can X take, it can take the value 2 if X takes the value 2 then what do you, how do you get the 2 2 back and say the chance that Y is equal to 0 given X is equal to 2. Let us call this step b. So, what happens here? So again, here, let us do a little, let me go and copy this, sorry for that, bring it down here so here what do you do? You say, no you toss a coin two times. And there are no heads and there are no heads and no head appears, and no head appears. That is one thing.

So, that means you know that this, you know, this is from the binomial probability, you know, that is the same as 1 minus P to the power of 2 with no head. Similarly, you can go and see chance that Y is equal to 1, given X equal to 2, that is the same as one head appears same as 2 choose 1, 2 choose 1, 2 choose 1 P power 1 and 1 minus P to the power 2 minus 1 and that is same as 2 into P into 1 minus P.

(Refer Slide Time: 10:35)

$$\mathbb{P}(Y=1 \mid X=1) = p \quad \ldots \text{(ie ... and outcome is a head)}$$

$$\mathbb{P}(Y=0 \mid X=2) = (1-p)^2 \quad \ldots \text{(ie toss a coin 2 times and no head appears)}$$

$$\mathbb{P}(Y=1 \mid X=2) = \binom{2}{1} p^1 (1-p)^{2-1} = 2\, p(1-p) \quad \ldots \text{(ie toss a coin 2 times and one head appears)}$$

$$\mathbb{P}(Y=2 \mid X=2) = p^2 \quad \ldots \text{(ie toss a coin 2 times and 2 heads appear)}$$

And that is the, that is like a going here again, and then let me just copy this and write it down again. Sorry about that. So, this is the same as toss a coin 2 times and 1 head appears So similarly, you can also have Y is equal to 2 as Y equals 2 given X equal to 2, that means you toss the coin twice you get 2 heads, that is the same as P square. So again, we go here, two heads appear.

So, that is how you would finish this computation. So, so in some sense, you are able to get the conditional distribution of Y given X. And the chance that you know what happens to X independently because X comes X occurs Y does not Y comes after X, so you can do X first that is easy to do and chance of Y given X values are the following.

(Refer Slide Time: 12:12)

(c) $\text{range}(Y) = \{0,1,2\}$

$$\begin{aligned}
\mathbb{P}(Y=0) &= \mathbb{P}((Y=0, X=1) \cup (Y=0, X=2)) \\
&= \mathbb{P}(Y=0, X=1) + \mathbb{P}(Y=0, X=2) \\
&= \underbrace{\mathbb{P}(Y=0 \mid X=1)}_{(b)} \underbrace{\mathbb{P}(X=1)}_{(a)} + \underbrace{\mathbb{P}(Y=0 \mid X=2)}_{(b)} \underbrace{\mathbb{P}(X=2)}_{(a)} \\
&= (1-p)\ \frac{1}{2} + (1-p)^2\ \frac{1}{2} \\
&= \frac{1}{2}(1-p) + (1-p)^2 .
\end{aligned}$$

So, how do you do the, how do you find distribution of Y, that is part C. In Part C, you find a distribution of Y. So, you want know that Y takes the value what, what range of Y is what? Range of Y is is 0, 1 or 2. That is the range of Y. Then the chance that Y is equal to 0 unified. So, now we know from the top Y can be 0 when X equal to 1 or 2.

That is the same as saying the chance that Y is equal to 0, X is equal to 1 union, let us call 0 and this union Y equal to 0, X equal to 2, they are disjoint events, same as chance that Y is equal to 0, X equal to 1 plus the chance that Y equal to 0, X equals to 2. And that is the same as Y equal to 0, this you can write as conditioned on X equal to 1 time the probability X equal to 1 plus the chance that Y equals to 0 conditioned on X equal to 2 times of probability X equals to 2.

So, now you go to step a and step b, you know all the answers. Step a has the answer that probability X equal to 0, 1 and 2 are half, and then conditional probability of Y equals to 0, X is equal to 1 is 1 minus P, and Y equals 0, X equal to 2 1 minus P squared. So, from a and b, you get the fact that this guy is just one half, this guy is just one half. That is just easy. This one half from a. And from b, this is just 1 minus P whole squared no heads to process and this just one heads. So, we will add the probability down that is the same as one half times 1 minus P plus 1 minus P the whole squared. That is some number. So, what would that be? That will just be hence leave it as it is.

(Refer Slide Time: 14:40)

$$\begin{aligned}
\mathbb{P}(Y=0) &= \mathbb{P}((Y=0, X=1) \cup (Y=0, X=2)) \\
&= \mathbb{P}(Y=0, X=1) + \mathbb{P}(Y=0, X=2) \\
&= \mathbb{P}(Y=0 \mid X=1)\,\mathbb{P}(X=1) + \mathbb{P}(Y=0 \mid X=2)\,\mathbb{P}(X=2) \\
&= \overset{(b)}{(1-p)}\;\overset{(a)}{\tfrac{1}{2}} + \overset{(b)}{(1-p)^2}\;\overset{(a)}{\tfrac{1}{2}} \\
&= \tfrac{1}{2}\left[(1-p) + (1-p)^2\right]
\end{aligned}$$

$\mathbb{P}(Y=1)$ . . . compute similarly

$\mathbb{P}(Y=2$

And then you could do the same thing, then you do the chance that Y is equal to 1. You can compute the same way, and compute similar. You conditioned on the outcome of Y in probability Y equals to 2, it will be same.

(Refer Slide Time: 15:21)

(c) $\text{range}(Y) = \{0,1,2\}$

$$\mathbb{P}(Y=0) = \mathbb{P}\big((Y=0, X=1) \cup (Y=0, X=2)\big)$$
$$= \mathbb{P}(Y=0, X=1) + \mathbb{P}(Y=0, X=2)$$
$$= \underbrace{\mathbb{P}(Y=0 \mid X=1)}_{(b)} \underbrace{\mathbb{P}(X=1)}_{(a)} + \underbrace{\mathbb{P}(Y=0 \mid X=2)}_{(b)} \underbrace{\mathbb{P}(X=2)}_{(a)}$$
$$= (1-p)\,\frac{1}{2} + (1-p)^2\,\frac{1}{2}$$
$$= \frac{1}{2}\left[(1-p) + (1-p)^2\right]$$

One can compute $\mathbb{P}(Y=1)$ & $\mathbb{P}(Y=2)$ similarly.

Now, one can compute probability Y equals to 1 and probability Y equals to 2 similarly. And we use the, so the idea was that somehow, we were not, we know that they are dependent, somehow X and Y are dependent to set it up properly, we first found the distribution of X, then we find the distribution of Y we had to find the conditional distribution of Y given X. And then once we did that, we were able to use that to find the distribution of Y. So, there were 3 steps to this problem. And there are 3 different notions involved.

(Refer Slide Time: 16:15)

✓ (a) $\mathbb{P}(X=1) = \frac{1}{2} = \mathbb{P}(X=2)$ as $X \sim \text{Uniform}\{1,2\}$

(b) $\mathbb{P}(Y=0 \mid X=1) = 1-p$ ... (ie toss a coin once and outcome is a Tail)

(Conditional Probability)

$\mathbb{P}(Y=1 \mid X=1) = p$ ... (ie toss a coin once and outcome is a Head)

$\mathbb{P}(Y=0 \mid X=2) = (1-p)^2$ ... (ie toss a coin 2 times and no head appears)

$\mathbb{P}(Y=1 \mid X=2) = \binom{2}{1} p^1 (1-p)^{2-1} = 2\,p(1-p)$ ... (ie toss a coin 2 times and one head appears)

So, now I will try and make these notions precise. So, one was the, this we got immediately, easily, this we used the idea to so these are to use the conditional probability idea, in this one.

(Refer Slide Time: 16:33)

(c) $\text{range}(Y) = \{0,1,2\}$

$$\mathbb{P}(Y=0) = \mathbb{P}((Y=0, X=1) \cup (Y=0, X=2))$$

Mutually exclusive event & (a), (b)

$$= \mathbb{P}(Y=0, X=1) + \mathbb{P}(Y=0, X=2)$$

$$= \mathbb{P}(Y=0 \mid X=1)\,\mathbb{P}(X=1) + \mathbb{P}(Y=0 \mid X=2)\,\mathbb{P}(X=2)$$

$$= (1-p)\,\frac{1}{2} + (1-p)^2\,\frac{1}{2}$$

$$= \frac{1}{2}\left[(1-p) + (1-p)^2\right]$$

One can compute $\mathbb{P}(Y=1)$ & $\mathbb{P}(Y=2)$ similarly.

And here are some how to find the distribution of Y, we had to use both that is the chance of that they were first split up over all possible outcomes of Y outcomes of X, and then you use conditional problem and a comma b is what you have. So, they all have they all have meanings in some sense, once you have structural random variables. So, let me write those definitions down. So, we have to go a little bit here to step back a little bit. So, I will view the random variables affected by an event rather than a variable, the random variable.

(Refer Slide Time: 17:20)

Definition 3.2.5 :- Let X be a random variable on a sample space S. Let $A \subseteq S$ be an event such that $\mathbb{P}(A) > 0$. Then the probability $\mathbb{Q}$ described by

$$\mathbb{Q}(B) = \mathbb{P}(X \in B \mid A) := \frac{\mathbb{P}((X \in B) \cap A)}{\mathbb{P}(A)}$$

is called the "conditional distribution" of X given the event A.

In previous example :- $A = \{X=1\}$ and we computed

$\mathbb{P}(Y=1 \mid A) = p$ and $\mathbb{P}(Y=0 \mid A) = 1-p$

So, let us start back a little bit. So, the definition 3.2.5. So here, what I do is I let X be a random variable on a sample space S and then A be an event such that probability of A was positive then the probability Q described by Q of B is the chance that X is in B given A. So, this is defined that this denotes the definition of chance that X and B and the event A divided by and the event A divided by the chance of probability of A, definition of that. This probability is called the conditional probability or the conditional distribution of X given data. That is what it is. So, given an event A I try and understand how experience.

So, the previous example, the previous example, so I notation was a little bit skewed because I used X for the first one and Y for the second one, you would call this as the as the conditional distribution of X or Y given X equal to 1. So, let me write it down in the previous example, you put the event A as the event X equals to 1. And we did, we computed, computed a probability of Y equal to 1, given A was P and probability of Y equal to 0, given A was equal to 1 minus P.

(Refer Slide Time: 20:44)

is called the "conditional distribution" of X given the event A.

In previous example:- $A = \{X=1\}$ and we computed

$\mathbb{P}(Y=1 \mid A) = p$ and $P(Y=0 \mid A) = 1-p$

$\therefore$ Conditional distribution $Y \mid A \sim$ Bernoulli (p)

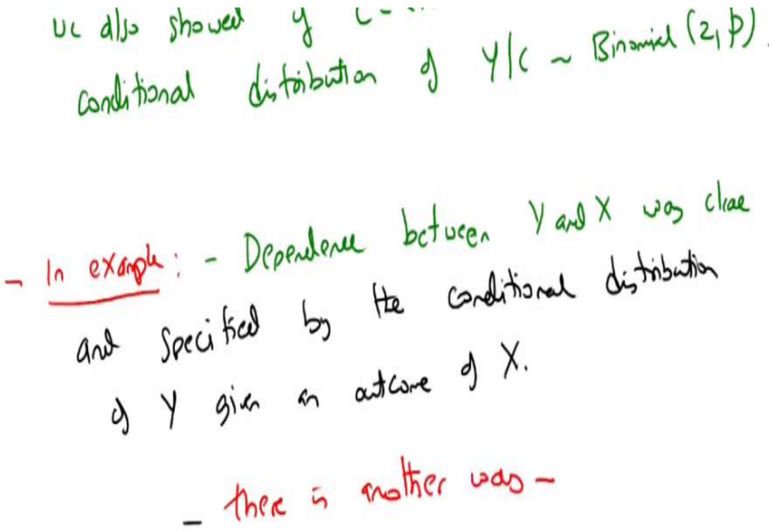
we also showed if $C = \{X=2\}$ then the conditional distribution of $Y \mid C \sim$ Binomial $(2, p)$.

Therefore, one could say that the conditional distribution of Y given A is Bernoulli. That is the sort of kind of idea described so far. And then also, if you look at if you just observe a little bit more, we also showed it let us say C was the event X equals to 2. Then we showed then, then then we showed the following, we also showed the following. So, that is the that the, the conditional distribution of Y given C was, if you look at carefully it was just binomial 2 comma P, that is what we shared.

So, the idea is that if you have an event or a random variable, that depends, you could classify first event consider random variable, and then understand the condition distribution of random variable given that event. So that is, that is the whole, so the the key to this whole thing is that the conditional probabilities are easy to determine in this above experiment.

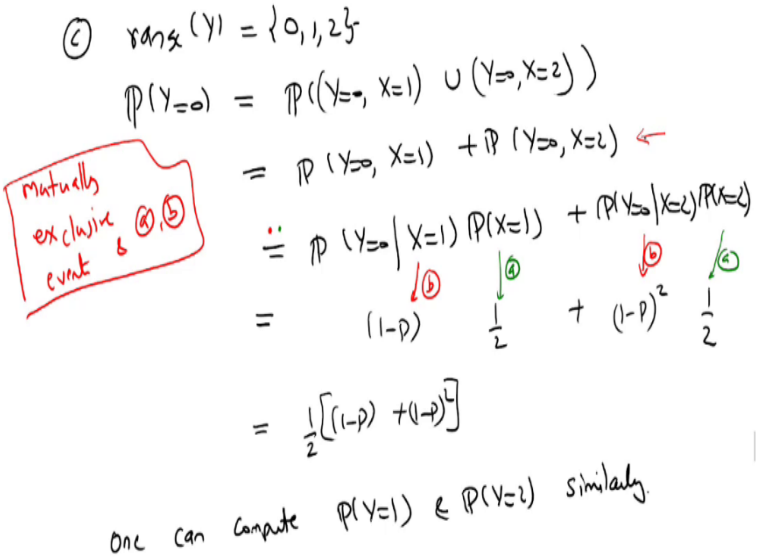
In the above experiment, we were able to determine this conditional probabilities. Because we knew exactly how to do that. Because we know that if X is equal to 1 you toss a coin once, then the two outcomes head or tail, the chance of each of them computed, then if you toss it twice, there are three possible outcomes, no heads, one head and two heads, and you know exactly how to compute. So, in some ways, the conditional probabilities were easy to determine, because Y was given in terms of X. That is one way of expressing different.

(Refer Slide Time: 23:27)



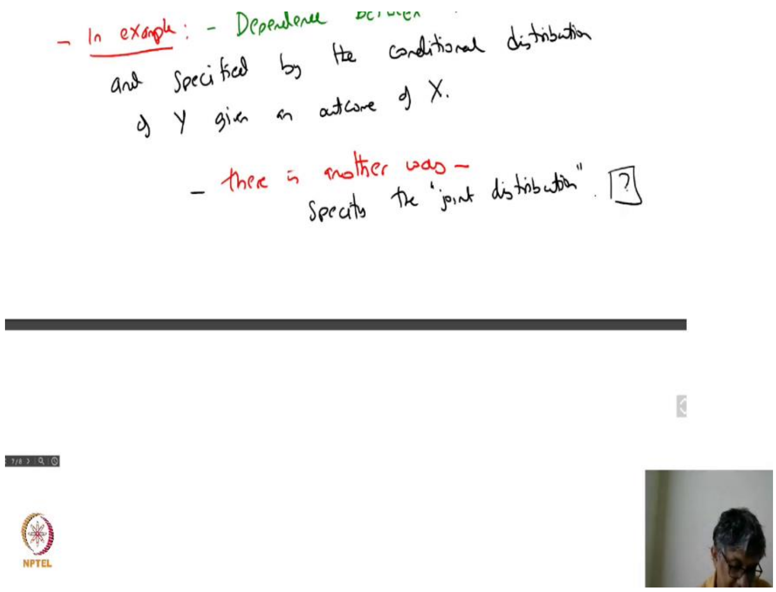
Now, so the let us see the background. So here in the example so dependence between X and Y between Y and X let us say was clear and more importantly, specified by the conditional distribution of Y given an outcome of X. That is the, that is one way we specified the events. There is another way to do this, there is another way there is another way, which I will describe in a second. The idea was that the idea is that so here we came from a to b, and then we came to c.

(Refer Slide Time: 24:40)



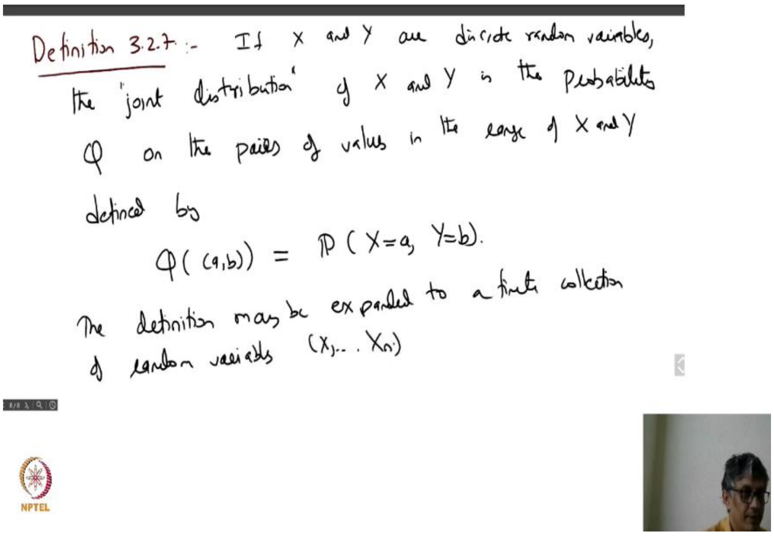
So, we could in fact use use, this is the step that we used in a and b were used here. So, in some sense, if I if I told you what this equality was, directly, then you could also go and find out the distribution of Y equal to P. So, that is another way of specifying dependence. So, which we will write down now. That is called the joint distribution.

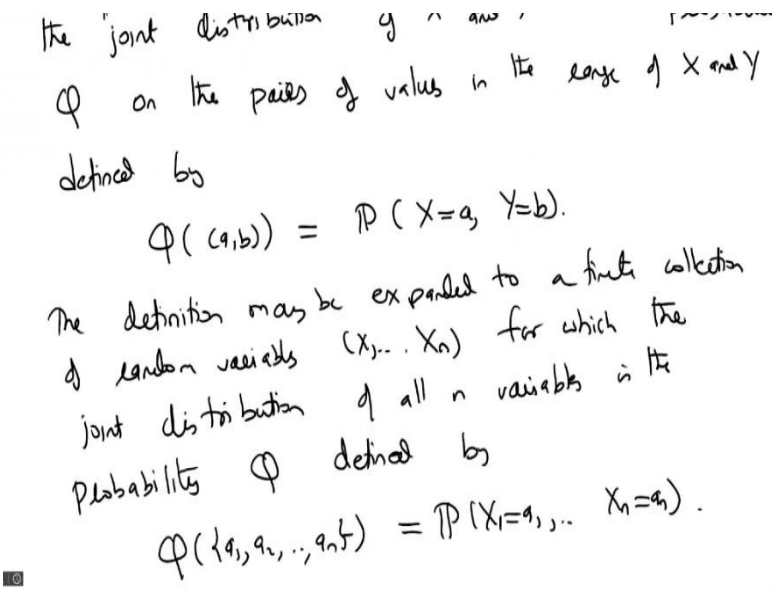
(Refer Slide Time: 25:15)



That is another way. So that is another way is you specify the specify sorry you specify the joint distribution. So, what do I mean by that? What do I mean by that? That is the following definition.

(Refer Slide Time: 25:50)



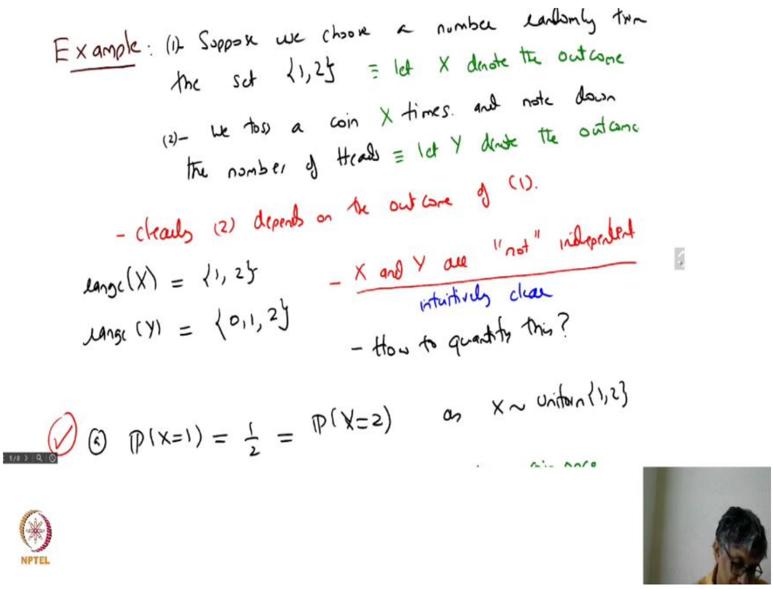


So, let us say definition, the definition is the following, if X and Y are two random variables. So, let us say all were discrete because I am not having here discrete random variables, random variables, in joint distribution of X and Y is the probability Q. Q is defined since crucial, Q is defined on the pairs of values probability Q on the pairs of values in the range of X and Y, defined by Q of a comma b is a pair is a chance that X is a and Y is b. That is the definition of joint distribution. You can also expand this definition to collection of n random variables, I will write that down. I will explain it next time.

So, the definition can also be expanded to a finite collection of random variables X1, X2, Xn for which the joint distribution of all n variables is the probability, probability, sorry why is this going away, is the probability Q defined by Q of the event a1, a2, an is the chance that X1 is a1, Xn is an.

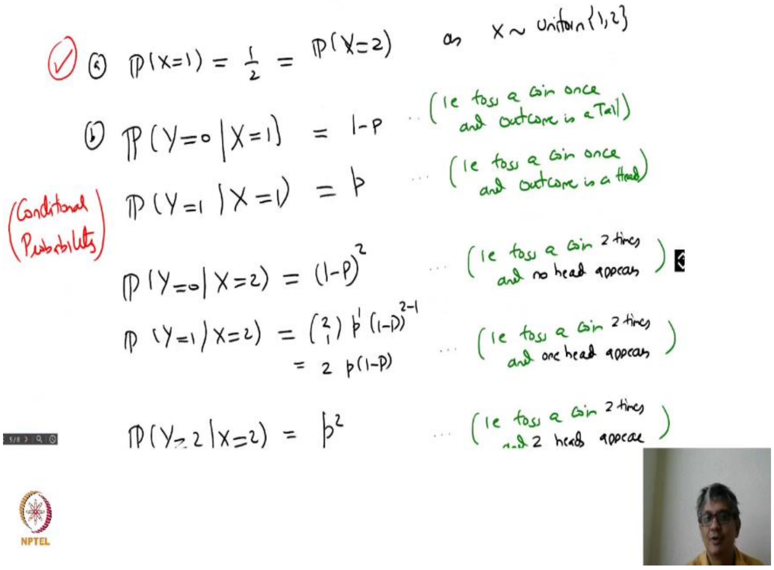
So, this is something that will build on the next class. So, the idea is that there is another way of specifying dependence. So, the idea is that, one way to do it is using the conditional distribution. The other way specifying the joint distribution of X and Y. One can already if you think a little bit one can sort of understand if, if the independent what happens, the dependent what happens to the joint law. We will discuss that next time. But as of now, I will just leave you with one thought that how we came to this idea.

(Refer Slide Time: 29:56)



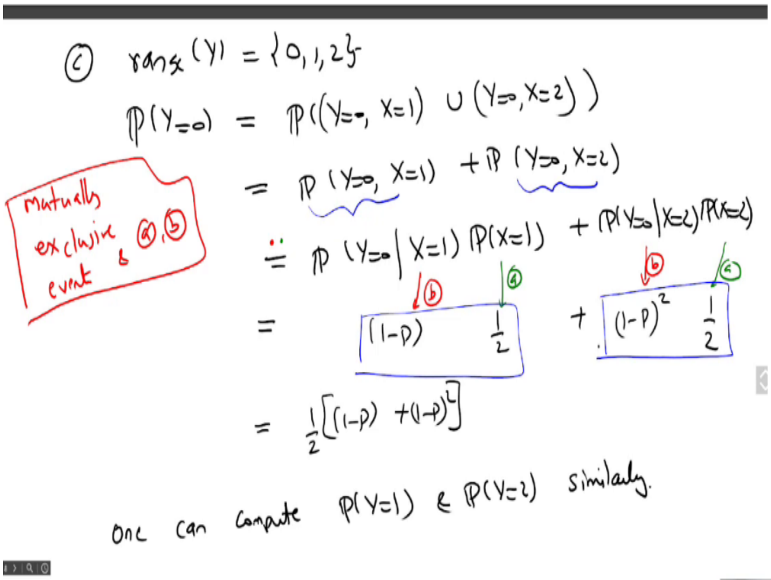
So, we took an experiment, well we chose a independent variable then we took a coin, we tossed it X times so that X is outcome of first experiment. So intuitively, it is clear that Y and X were dependent.

(Refer Slide Time: 30:14)



Now one, how we put calculation, one was, it was easy to compute the distribution of X it was easy because X is uniform 1. But for Y, the way we did it was we specified the conditional law of X or Y given X. That is an experiment because we could understand how the experiment was specified. And the probability was able to compute.

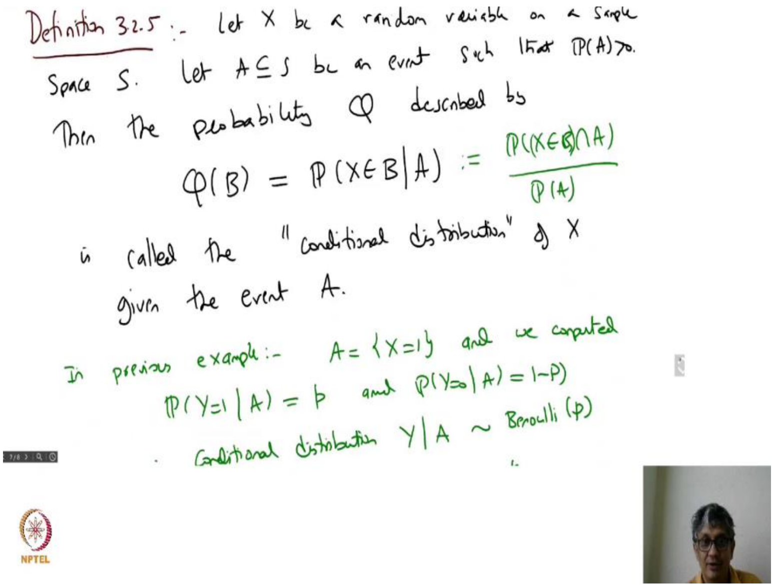
(Refer Slide Time: 30:39)



Now, from this one, we had to go and compute the chance that Y equals 0. And also, here, you are able to compute the joint law. So, the example of these two computations let me write this in blue so these two computations from here to here, you come all the way here. Sorry not this one. I will delete this.

So, this specified the joint distribution. And that computation is hidden in this calculation. This is the joint distribution of Y equal to 0, joint probability of Y equals 0 and X equals to 1. And this was the joint probability of Y… This whole thing, this was the probability of Y equals 0, X equals to 1, this probability was this time this with a probability of Y equals to 0.

(Refer Slide Time: 31:30)



So, there were two ways of understanding dependence one was by the conditional law, and the other was by the by specifying the joint distribution. So, we see examples of how to do this and how independence come into this…