PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 04

Lecture - 18

Binomial Distribution

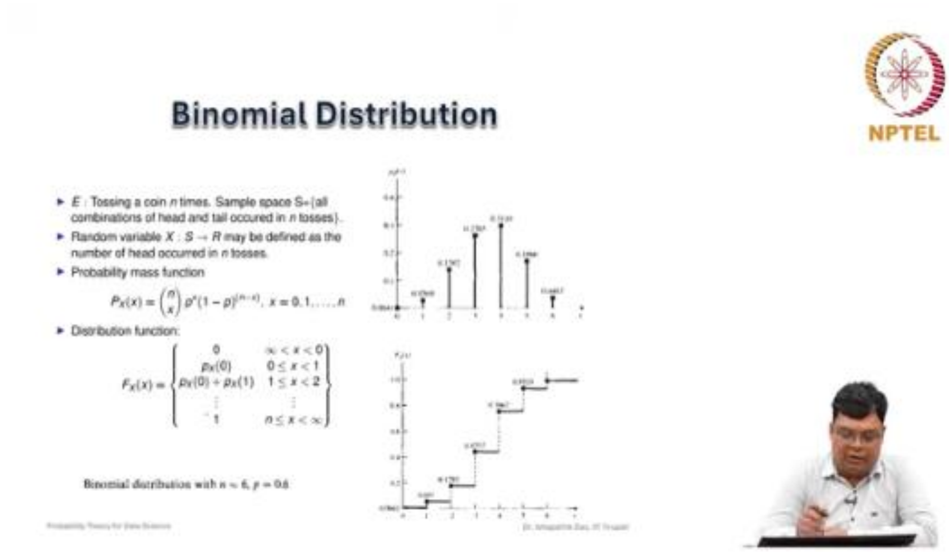
Now, we will discuss the binomial distribution function. So, the binomial distribution function is, suppose you are tossing a coin not only once; it is suppose n number of times you are tossing a coin. This coin has a fixed probability of heads, which is p. Then, suppose you are tossing the coin 10 times. How many successes can you get?

It can be 0, it can be 1, or it can be a maximum of 10 successes, meaning success refers to getting heads. In the real world, whenever in a clinical trial, we frequently come across this kind of data. Suppose you are giving medicine to n patients, the same medicine, the same dose, and you want to see what the success rate is. Suppose you are giving the vaccine and you want to see how it is working. There are n patients, and because it is random, you may get success or may not get success.

There is a success probability, p, fixed for each trial. So, then out of n patients, how many of them will be cured by that medicine, or how well the vaccine will work for the n number of patients? In general, you are tossing a coin n times, and the sample space S consists of all combinations of heads and tails that occur in n tosses. This is the sample space whenever you are tossing a coin n times. So now, the random variable X can be defined as the number of heads that occurred in n tosses, which means the number of successes in n tosses.

Then, how do you derive this probability mass function? So let us see that. Suppose if you see here whenever we are considering n number of times, suppose let us consider n =

3 to understand clearly. So you have tossed the coin 3 times. Then what will be the possibilities?



Suppose 3 times you are tossing the coin and every time the toss gives an observation: x_1 is the observation that first time you toss the coin, x_2 is the observation the second time you toss the coin, and x_3 is the observation the third time you toss the coin. So this value may be 0 or 1; this is also 1 or 0. So if you take this random variable $y = x_1 + x_2 + x_3$, then what it is giving? This is nothing but 3 times you are tossing the coin, and suppose there are 2 successes; then y = 2. If there are 3 successes, then y = 3. If there are 0, then it is giving the actual number of successes.

Now, how can you find the probability of the number of successes? Suppose you want to find this random variable probability that y = 0. So y = 0 can be possible because the first time you are tossing the coin this also will be 0. The second time you are tossing the coin, this observation also will be tail (0), and then the third time also (0), then that is the only possibility. So then what is the probability?

We are assuming that all x_i are from the previous distribution, which is called the Bernoulli distribution. We denote this Bernoulli distribution; we did not mention it. We denote it like this: Bernoulli(p). So, Bernoulli(p) means we know that $P(x_i = 1) = p$, and the probability that $x_i = 0$ will be equal to 1 - p. Since you are fixing the coin only, then all have the same distribution.

So similarly, you can just replace it by i for i = 1, 2, and 3. All have the same distribution because you are fixing the coin and it has the same probability. The first time you toss the coin, what is the probability of heads? The second time you are tossing the coin, what is the probability of heads? It will be the same, so that is why the first time this observation is 0.

What is the probability? This is nothing but 1 - p. The second time you are tossing the coin, what is the probability that it will be again 0? This is 1 - p. The third time also, it is again 0; this is also a probability of 1 - p.

Because they are independently tossed, you know that whenever events are independent, then this probability will be multiplied. So that is why this probability will be (1 - p) × (1 - p) × (1 - p). Hence, $P(y = 0) = (1 - p)^3$. So this is the probability.

Now, first of all, you can understand what will be the range of this random variable. The range of y will be nothing but the number of successes. It may be the minimum value of 0; it can be 1 also, and it can be 2 also; it can be 3 also. So that means the range will be how many times you are tossing the coin. So then it will be 0, 1, 2, 3. If n = 4, then it will be 0, 1, 2, 3, 4. So then you have to find the probability of y = 0; we found that.

So then we have to find the probability of y = 1 because we want to find the probability mass function. So the probability of y = 1 is how it can be possible. Note that y = 1 means that the sum of the values will be 1, so that is why if $x_1 = 1$, then the other two will be 0. So this is nothing but the possibility. One possibility is this: the first time it is 1, but this is not the only possibility; the next two will be 0.

Another possibility will be (1, 0, 0), and another possibility will be (0, 0, 1). In all these cases, the sum is y = 1, and they are mutually exclusive events because the intersection of this event with this event is equal to the null set. So what is the probability that it may appear? Because it is one probability, the probability that $x_1 = 1$; this is nothing but p, and $x_2 = 0$; the second observation is 0; this probability is 1 - p, and the third observation is again 0; this is 1 - p. So this is nothing but $p \times (1 - p)^2$.

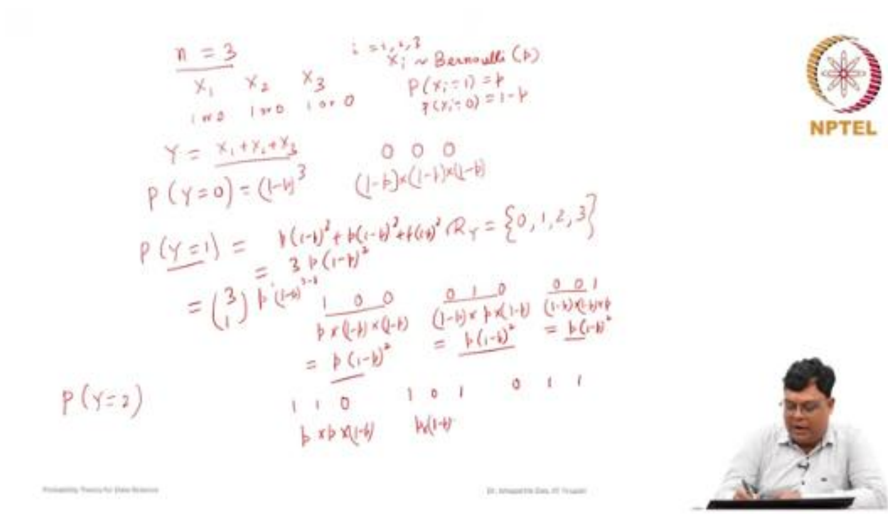
Now, what will this case be? In this case, the first observation probability that $x_1 = 0$; this probability is 1 - p, then multiplied by the probability that $x_2 = 1$; this is p, and the probability that $x_3 = 0$; this is 1 - p. So in that case, this is also $p \times (1 - p)^2$. Similarly, here you can see that this is (1 - p) × (1 - p) × p. So this is nothing but $p \times (1 - p)^2$.

Now, because they are mutually exclusive, the possibility of y = 1 is nothing but the probability of all these observations. So it is the union of these mutually exclusive events, and the probability of the union of this event will be nothing but the sum of their individual probabilities. Since they are all the same, it is nothing but $p \times (1 - p)^2 + p \times (1 - p)^2 + p \times (1 - p)^2$. So, this is nothing but $3 \times p \times (1 - p)^2$.

Now, we want to understand the pattern of the probability so that we can write it in a general form.

So what is the pattern? The pattern is that we want success 1 out of a total number of 3. That means how many mutually exclusive events can we say we can find in the place where one position will be 1 and the remaining will be 0. That many combinations we can find. Every time that 1 is p, and the remaining (1 - p) will be multiplied, so this way it is $p^1 \times (1 - p)^{3-1}$, and 3 - 1 will be the remaining 1 - p.

So that is why this is nothing but 3 × p × (1 - p). Similarly, if you can find, you can see that the probability of y = 2. How we can find that 2 means that 2 positions will be 1. That many (3 choose 2) positions, and every time this probability will be p × p × (1 - p). Similarly, here also, 2p × (1 - p) × p, and similarly here, (1 - p) × p × p.



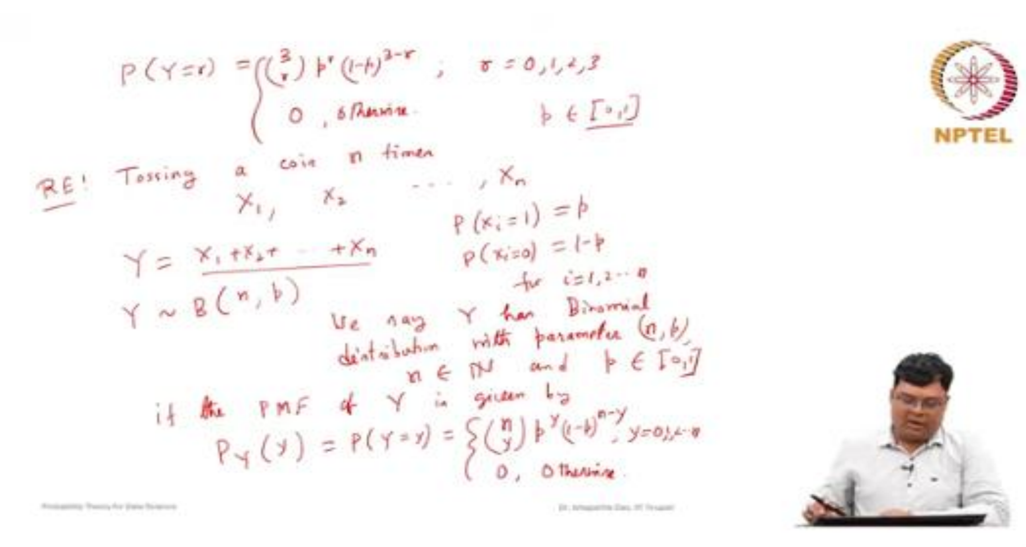
So, all these are the same: $p^2 \times (1 - p)$. This is $p^2 \times (1 - p)$, and $p^2 \times (1 - p)$. So that is why this probability is nothing but $3 \times p^2 \times (1 - p)$. So, it can also be written as ${}^3C_2 \times p^2 \times (1 - p)^{(3 - 2)}$. In general, we can write this probability that Y = r, this is nothing but ${}^3C^r \times p^r \times (1 - p)^{(3 - r)}$. Whenever r can be 0, 1, 2, or 3, it is 0 otherwise.

So, now we have tossed this coin only 3 times, n = 3 here. So, in the general case, suppose we are tossing a coin. So the random experiment is nothing but tossing a coin n times. Tossing a coin n times and then your observation will be, first time this is X_1, then second time X_2, like that you have Xn. We say that they are identically distributed because all have the same distribution, they are all Bernoulli distribution. $X_i = 1$ is p and the probability that $X_i = 0$, this is 1 - p, for i = 1 to n.

And they are also independent because you are tossing a coin n times. So, the first observation does not depend upon the second observation. They are independently distributed random variables. Now, if you consider this $Y = X_1 + X_2 + ... + Xn$. So, that is nothing but $\sum X_i$.

Then Y has, we say that Y has binomial distribution, with two parameters here. So, this p is a parameter because this distribution depends upon this value of the probability of success. Here, this distribution depends upon two parameters, two values. How many times you are tossing the coin? That is n. And what is the probability of success, or the probability of heads of the coin?

So, Y has a binomial distribution with parameters n, p, where n is a natural number because it can be 1, 2, 3, 4, and p is the probability of success. So, Y has a binomial distribution with parameters n, p, and if the probability mass function of Y is given by, see that how we will generalize it. So, here three times we tossed the coin. So, then this probability mass function looks like this: $P(Y = y) = {}^{n}C_{\gamma} \times p^{Y} \times (1 - p)$^$(n - y)$, whenever y = 0, 1, up to n; otherwise, it is equal to 0. This is the probability mass function of a binomial distribution.



Suppose we have tossed the n number of times, and then p is the probability associated with the event. The probability mass function is given by P(y), which represents the probability that Y takes the values y and is expressed as:

$P(Y = y) = nCy \times p^y \times (1 - p)^{(n - y)}$ for $y = 0, 1, 2, \ldots, n$; it equals 0 otherwise. Now, let us consider how we can represent it graphically. First, what is the probability of $Y = 0$? This is calculated as:

$P(Y = 0) = nC0 \times p^0 \times (1 - p)^n = (1 - p)^n$, which we denote as P0.

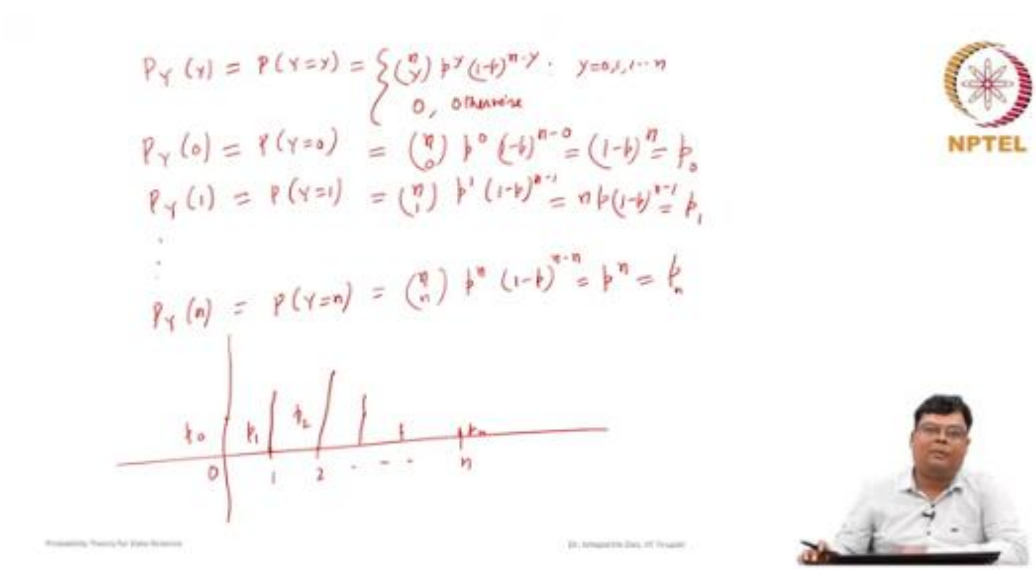
Next, what is $P(Y = 1)$? This is the probability that $Y = 1$, calculated as:

$P(Y = 1) = nC1 \times p^1 \times (1 - p)^{(n - 1)} = n \times p \times (1 - p)^{(n - 1)}$, which we denote as P1. For $Y = n$, we have:

$P(Y = n) = nCn \times p^n \times (1 - p)^0 = p^n$, which we denote as Pn.

If we represent this graphically, we will have values at $Y = 0$, $Y = 1$, and so on, up to $Y = n$. The appearance of the graph depends on the value of p. Typically, it may start high or low at $Y = 0$ and change shape as p varies. After some values, the graph will decrease, depending on the values of p. This is the probability mass function.

Now, suppose you want to find the cumulative distribution function from this probability mass function. The CDF, or cumulative distribution function of X, is given by—suppose we want to find the cumulative distribution function of Y. Sorry, we denote this by Y, the binomial distribution, and we denote Y by F(Y).

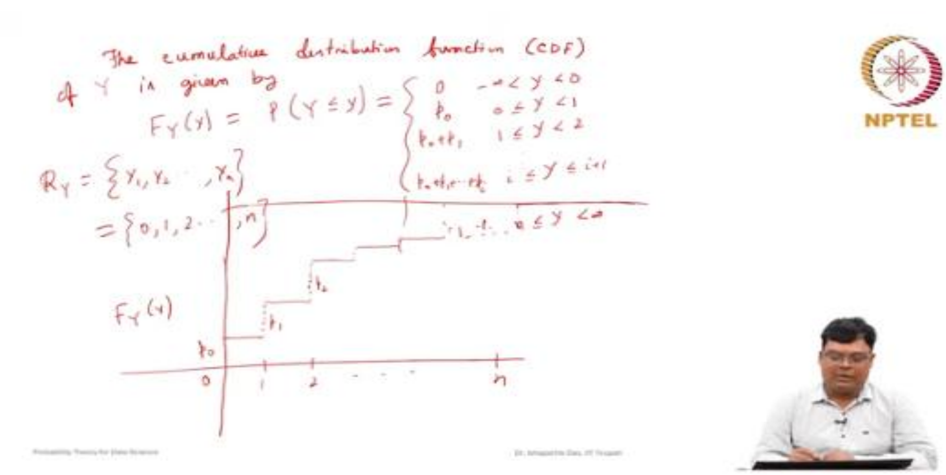


This is simply the probability that $Y \leq y$. We know the probability mass function here, and we also learned how to find the cumulative distribution function from the probability mass function. So, how do we find it? What is the range of this random variable Y? Usually, if you write y_1, y_2, up to y_n, this range is just 0, 1, 2, ..., n.

So, this is the range—the minimum value is 0, then 1, and so on. Whenever $Y < 0$, there is no probability mass, so the probability will be 0 when $Y < 0$. Now, whenever $Y \geq 0$ and < 1, it will simply be P_0 at 0. For the next interval, from 1 to < 2, it will be $P_0 + P_1$, and so on. This is something we have already discussed when talking about the probability mass function of a discrete random variable and how to find the cumulative distribution function from the probability mass function.

Similarly, for any interval greater than or equal to i and less than or equal to i + 1, where $i + 1 \leq n$, this is nothing but $P_0 + P_1 + ... + P_i$. In this way, whenever $Y \geq n$ and $< \infty$, this will be equal to 1. So, we have found the cumulative distribution function of the binomial distribution. If you want to draw the graph of the cumulative distribution function, you'll see that the values are 0, 1, 2, ..., n. For example, up to 0, $F(Y) = 0$.

At 0, it's P_0; at 1, it's $P_0 + P_1$; at 2, it's $P_0 + P_1 + P_2$, and so on. It continues in this way, converging. After n, it will be 1. So, this will be the cumulative distribution function up to n, and beyond that, it remains equal to 1. So, this is the graph of the cumulative distribution function of the binomial random variable. Now, we'll find the mean and moments of the binomial random variable.



Let's start with the mean. What will the mean be? The first-order raw moment, denoted as μ_1 (or the expected value of Y). By definition, this is simply the summation of all values it takes, each multiplied by the probability of that value, which is the sum of y_k times the probability of $Y = y_k$. So, y_k takes values from 0 to n. Therefore, this becomes:

Summation from k = 0 to n of k × P(Y = k)

Since the term for k = 0 is zero, we can omit it. So, we start from k = 1 to n, summing k times the probability of Y = k. Now, what is the probability mass function here? We already defined it as:

P(Y = k) = nCk × p^k × (1 - p)^(n - k)

Now, we want to simplify this expression. It becomes:

Summation from k = 1 to n of k × nCk × p^k × (1 - p)^(n - k)

We can rewrite nCk as:

nCk = n! / (k! × (n - k)!)

Thus, the sum becomes:

Summation from k = 1 to n of k × (n! / (k! × (n - k)!)) × p^k × (1 - p)^(n - k)

Now, we simplify further. We write n! = n × (n - 1)!, so we can factor out n from the sum:

n × Summation from k = 1 to n of ((n - 1)! / ((k - 1)! × (n - k)!)) × p^k × (1 - p)^(n - k)

We can further rewrite p^k as:

p × p^(k - 1)

Now, the expression becomes:

n × p × Summation from k = 1 to n of ((n - 1)! / ((k - 1)! × (n - k)!)) × p^(k - 1) × (1 - p)^(n - k)

To simplify further, let r = k - 1 and m = n - 1, assuming n ≥ 1. If n = 1, then m = 0, which corresponds to the Bernoulli distribution, a special case of the binomial distribution. So, assuming n > 1, we get the general binomial distribution. Now, with this transformation, as k goes from 1 to n, r will go from 0 to m = n - 1. Thus, (n - 1)!

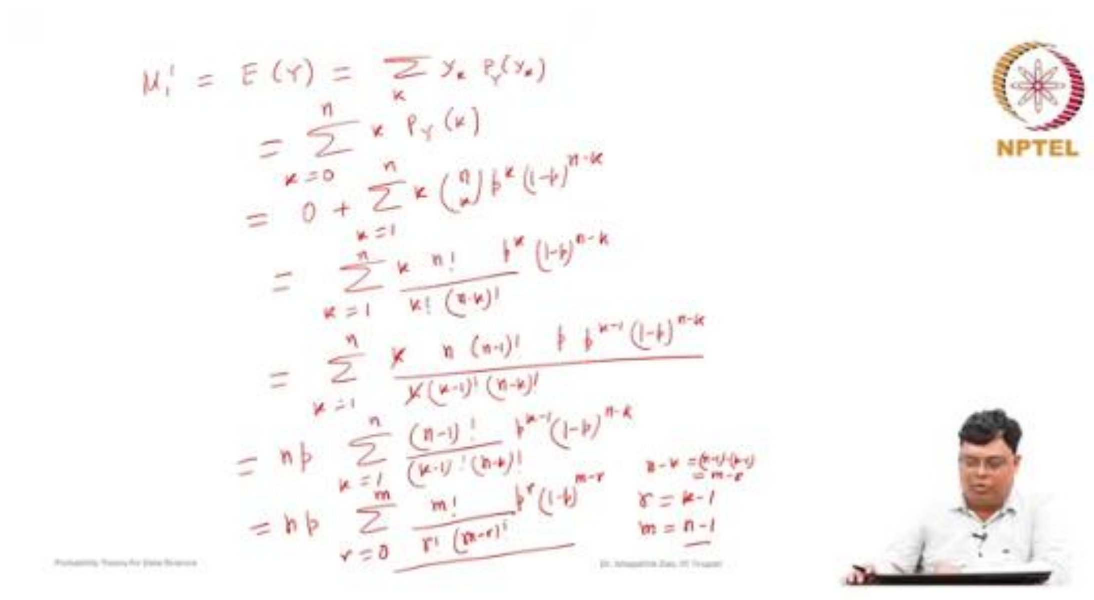
becomes m!, and the limits of r are from 0 to m. We can rewrite this in terms of m and r, with factorial terms like m!, r!, and (m - r)!, along with $p^r \times (1 - p)^{(m - r)}$.

This expression now resembles a known formula, which can help us simplify further. So, what is that known formula? Suppose you have a binomial expansion. You know that for $(a + b)^2$, this is nothing but:

$a^2 + 2ab + b^2$

Now, in general, if you write $(a + b)^m$, what is the formula? This is a binomial expansion, which is simply:

Summation from r = 0 to m of (mCr) × $a^r \times b^{(m - r)}$



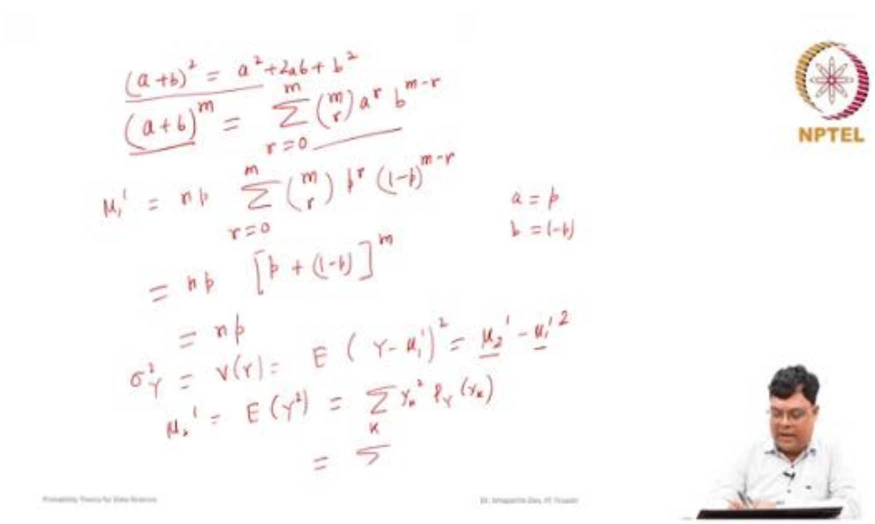
If you set m = 2, you get $(a + b)^2$, and if m = 3, you get the expansion for $(a + b)^3$. This is a general formula for any m ≥ 2. Now, here's what we found: we found that μ_1 is nothing but n × p × Σ from r = 0 to m. You can see here that we have m! / (r! × (m - r)!), which is just mCr. Then we have $p^r \times (1 - p)^{(m - r)}$.

If you compare this with the binomial expansion, setting a = p and b = (1 - p), you find that this is nothing but n × p × $(a + b)^m$, which simplifies to n × p × 1^m, or simply n × p. So, finally, we got the expected value of Y.

Now, let's move to finding the variance of the binomial distribution. We've seen how to find the mean, which is simply $n \times p$. This is useful to remember because whenever we say Y follows a binomial distribution with parameters n and p, the expected value of Y is known to be $n \times p$.

Next, if we want to find the variance, we'll use the formula for variance. We know that the variance of Y, denoted as σ^2_γ, is the expected value of $Y^2 - (\mu_\gamma)^2$, where μ_γ is μ_1 as we found here. In simplified terms, we use $\mu_2 - \mu_1^2$. Since we know $\mu_1 = n \times p$, now we need to find μ_2, which is the expected value of Y^2. The expected value of Y^2 is nothing but the summation of $y_k^2 \times P(Y = y_k)$. Here, y_k takes values from 0 to n.

So, we can represent this as Σ from $k = 0$ to n of $y_k^2 \times P(Y = y_k)$. So now, if we represent this, the variance of Y, σ^2_γ, what we will find? This is nothing but the variance of Y. So the variance of Y will be Σ from $k = 0$ to n of $k^2 \times P(Y = k)$.



First, we are trying to find μ_2. So, μ_2 is nothing but the expected value of Y^2. That we wrote here. This is nothing but:

$\sum$ (from $k = 0$ to n) of $k^2 \times P(Y = k)$

Now, if you write this, you can see that when $k = 0$, this value is 0. So we have:

$0 + \sum$ (from $k = 1$ to n) of $k^2 \times P(Y = k)$

where:

$P(Y = k) = nCk \times p^k \times (1 - p)^{(n - k)}$

Now, if you use the same technique as earlier, you can see that:

$k^2 \times (n! / (k! \times (n - k)!))$

is nothing but $P(Y = k)$. Now you can see here that k ranges from 1 to n. This k^2 is actually:

$n! = n \times (n - 1)!$

We can write k as:

$k \times (k - 1)!$

and this is:

$(n - k)!$

Also, p^k we can write as:

$p \times p^{(k - 1)}$

and this is:

$(1 - p)^{(n - k)}$

Now here we can cancel out terms, but there is still another k, so we cannot cancel this k as we did earlier. Hence, we cannot take the similar type of transformation, and we cannot find a closed form.

So next, we will use a different technique to find μ_2. Instead of finding the expected value of Y^2, we will find the expected value of $Y \times (Y - 1)$, because you can see that k^2, we cannot cancel. However, we can cancel $k \times (k - 1)$. It can be further written as:

$k \times (k - 1) \times (k - 2)!$

So, we will find the expected value of $Y \times (Y - 1)$. If we can find it, then we can express it as:

$Y^2 - Y$

Thus, this is:

$E(Y^2) - E(Y)$

Now, the expected value of Y is:

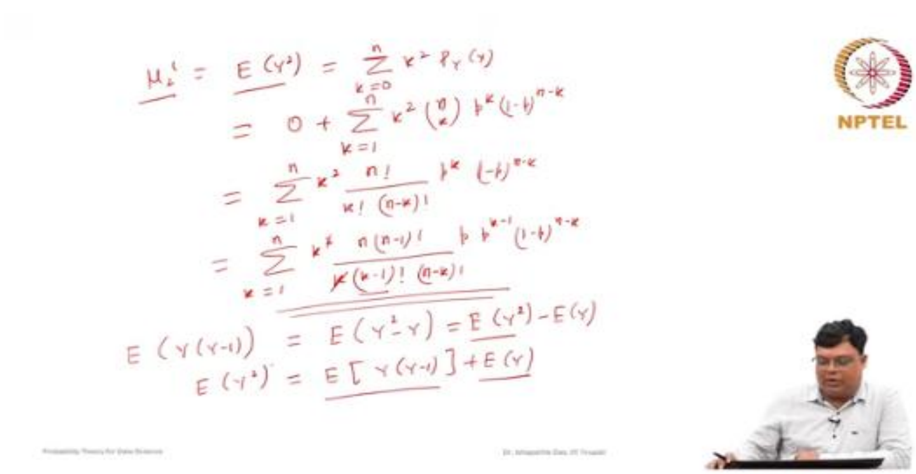
$E(Y) = n \times p$

which we have already computed earlier. Now, we need to compute the expected value of $Y \times (Y - 1)$. So, the expected value of $Y \times (Y - 1)$ is defined as:

$E(Y \times (Y - 1)) = \sum$ (from k = 0 to n) of $k \times (k - 1) \times P(Y = k)$

which is:

$\sum$ (from k = 0 to n) of $k \times (k - 1)$



We have already written this as n choose k. So, we have:

$nCk \times p^k \times (1 - p)^{(n - k)}$

Now, in this summation, whenever you put k = 0, the value is 0. Also, when k = 1, the value is 0. So, we have the summation from k = 2 to n. Assuming that $n \geq 2$, we have:

$k \times (k - 1)$

Now, we can write n choose k as:

$n! / (k! \times (n - k)!)$

which is multiplied by:

$p^k \times (1 - p)^{(n - k)}$

This is equal to:

$\sum$ (from $k = 2$ to n) of $k \times (k - 1)$

We can write n! as:

$n \times (n - 1) \times (n - 2)!$

and k! as:

$k \times (k - 1) \times (k - 2)!$

The remaining term will be $(n - k)!$, and we can write p as:

$p^2 \times p^{(k - 2)}$

while $(1 - p)$ remains as it is.

Now, we can cancel k and $(k - 1)$ because k ranges from 2 to n, so we are not dealing with 0/0. Thus, we can cancel them since they are greater than 0. Now, $n \times p^2$ is independent of k, so it can be taken outside the sum, resulting in:

$n \times (n - 1) \times p^2$

The remaining sum will be:

$\sum$ (from $k = 2$ to n) of $(n - 2)! / ((k - 2)! \times (n - k)!) \times p^{(k - 2)} \times (1 - p)^{(n - k)}$

Now, we will take a transformation similar to the one we used earlier. We will take:

$r = k - 2$ and $m = n - 2$

You can see that the limit for r starts at $k = 2$, so the minimum limit for r is 0. When $k = n$, $r = n - 2$. Thus, we have r going from 0 to m. $(n - 2)!$ is nothing but:

$m!$

divided by $(k - 2)$, which is r!, and $(n - k)$ is equal to $(n - 2) - k + 2$, which we can represent as:

$m - r$

This gives us:

(m - r)!

Then, we have:

p^p and $(1 - p)^{(m - r)}$

As we did previously, you can see that from this binomial expansion, if you consider a = p and b = (1 - p), this results in:

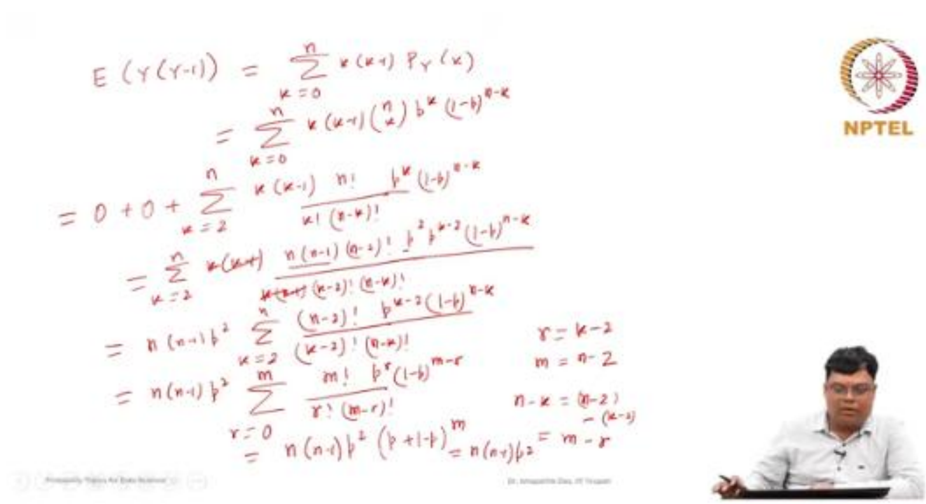
$n \times (n - 1) \times p^2$

This is equal to (p + 1 - p) raised to the power of m. Since p + 1 - p = 1, this simplifies to:

$n \times (n - 1) \times p^2$

So, finally, we found that the expected value of Y × (Y - 1) is:

$n \times (n - 1) \times p^2$



Now, we have μ_2, which represents the expected value of Y^2. This value is:

$n \times (n - 1) \times p^2 + n \times p$.

We will simplify this. μ_2 is the expected value of Y^2, which is:

$n \times (n - 1) \times p^2 + n \times p$.

Now, what will be the variance?

The variance of Y, σ^2_γ, is μ_2. We know that the formula for variance is:

$\sigma^2_\gamma = E(Y^2) - (E(Y))^2$.

The simplified formula is:

$\sigma^2_\gamma = \mu_2 - \mu_1^2$.

We found μ_2 to be:

$n \times (n - 1) \times p^2 + n \times p - n \times p^2$,

which is μ_1. Now, if we simplify this, we get:

$n^2 \times p^2 - n \times p^2 + n \times p - n^2 \times p^2$.

The $n^2 \times p^2$ terms will cancel. If we take $n \times p$ as common, we get:

$n \times p \times (1 - p)$.

We say that this is:

$n \times p \times q$,

where $q = 1 - p$.

So we found that the expected value of Y, which is the mean of Y, is:

$\mu_1 = n \times p$,

and the variance of Y, σ^2_γ, is:

$\sigma^2_\gamma = n \times p \times q$,

where $q = 1 - p$.

So these are some basic properties of the binomial distribution that we discussed.

We found the probability mass function and graphically represented it, and we also discussed how the cumulative distribution function looks; it will be a step function. We found that the mean of the binomial distribution is:

$E(Y) = n \times p$,

and the variance of the binomial distribution is:

$\sigma^2_{\gamma} = n \times p \times q$.

So, next, we will discuss some more important distributions like Poisson distributions and other continuous distributions.