PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 04

Lecture - 17

Numerical Examples of Moments and Bernoulli Distribution

You can remember that for this discrete random variable, let X be a discrete random variable with the probability mass function given by $P(X) = 1/4$ whenever $X = 1$ or $X = 2$, and $P(X) = 1/2$ whenever $X = 3$. This is equal to 0 otherwise. You can check that this is a probability mass function because it is always greater than or equal to 0, and if you take the sum of all the values, it equals 1. However, it is not a uniform distribution since the probabilities are different; for $X = 3$, it is 1/2, and for $X = 1$ and $X = 2$, it is 1/4.

Now, what will be the mean and variance?

To find the mean, the expected value, denoted as μ_X, is defined as the summation $\sum x_k P(X_k)$. The values that X can take are 1, 2, and 3. Thus, the expected value can be calculated as $E(X) = 1 \times P(X=1) + 2 \times P(X=2) + 3 \times P(X=3)$. Substituting the probabilities gives us

$$E(X) = 1 \times (1/4) + 2 \times (1/4) + 3 \times (1/2).$$

This simplifies to

$$E(X) = (1/4) + (2/4) + (6/4) = 9/4.$$

Now, this is simply a notation, where $\mu_X = 9/4$. If you want to find the variance, the formula is given by σ^2_X, which is the expected value of $(X - \mu_X)^2$. This can be computed as the summation $\sum (x_k - \mu_X)^2 P(X_k)$. To find this, you will need to perform the subtraction, square it, and then multiply by the respective probabilities. However, instead of calculating it directly, we can use another formula to find the variance.

What we will do is find the second-order raw moment. The second-order raw moment is nothing but μ'_2, which is the expected value of X^2. So, the expected value of X^2 is calculated as

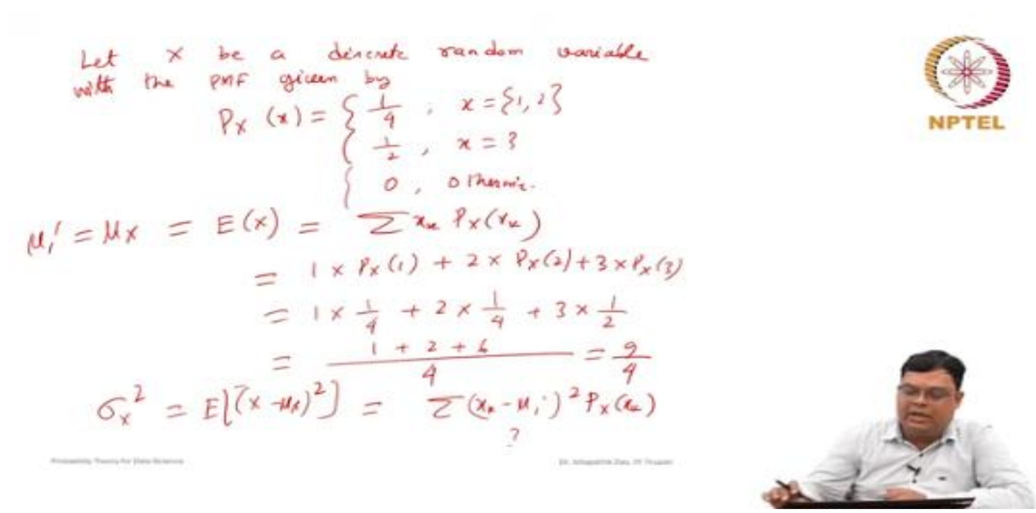
$$E(X^2) = \sum (X_k^2 \times P(X = X_k)),$$

where the sum is over all possible values that X can take. We have

$$E(X^2) = 1^2 \times P(X = 1) + 2^2 \times P(X = 2) + 3^2 \times P(X = 3) = 1 \times (1/4) + 4 \times (1/4) + 9 \times (1/2).$$

Calculating this gives us

$$E(X^2) = (1/4) + (1) + (9/2) = (1/4) + (4/4) + (18/4) = 23/4.$$



Next, the variance of X, denoted as σ^2_X, is calculated using the formula $\sigma^2_X = \mu'_2 - (\mu'_1)^2$. We already found that $\mu'_2 = 23/4$ and $\mu'_1 = 9/4$. So, we compute

$$\sigma^2_X = 23/4 - (9/4)^2.$$

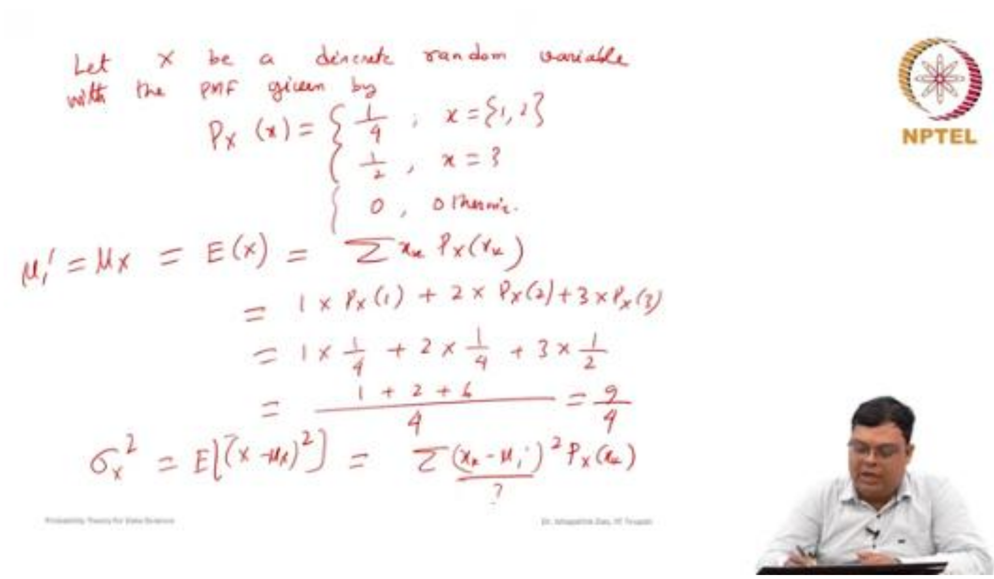
Calculating $(9/4)^2$ gives us 81/16. Now, substituting this back into the variance formula gives us

$$\sigma^2_X = 23/4 - 81/16.$$

To simplify this, we need a common denominator. The common denominator between 4 and 16 is 16. Thus, $23/4 = 92/16$, so substituting gives us

$\sigma^2_X = 92/16 - 81/16 = 11/16$.

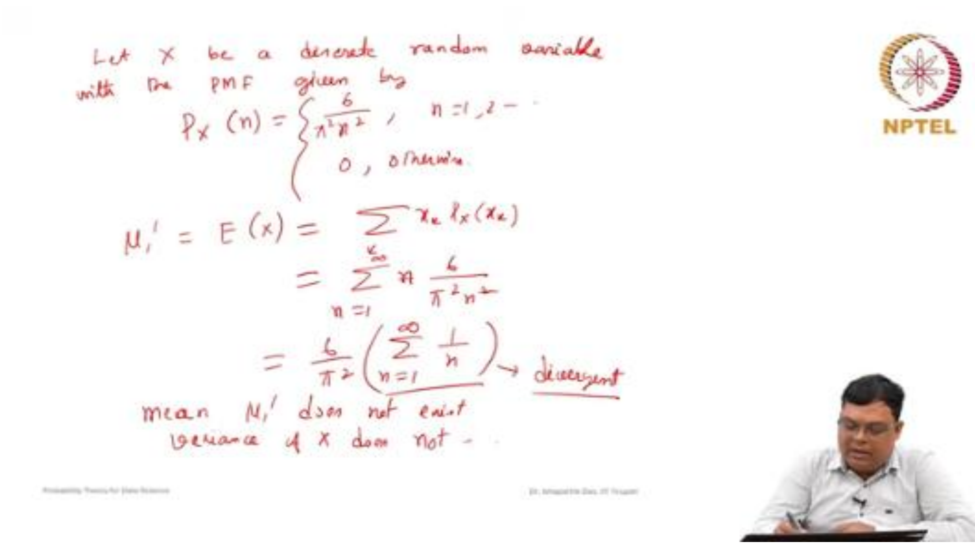
So, this variance can be easily found, but you can also try calculating it directly from the definition to see if you get the same value; it may be a bit more complicated. You can check whether you are getting the same value because we have simply used the actual definition, and then we simplified this formula. Now, before moving to another example, we will discuss continuous random variable cases.



Here, we will discuss another discrete random variable example. Remember, this is for the infinite case, and we will work through more examples. For the continuous case and discrete case, you can recall that the example given is $P(X = n) = 6/\pi^2 n^2$ for $n = 1$ to infinity and 0 otherwise. Let us consider this example where X is a discrete random variable with the probability mass function given by $P(X = x) = 6/\pi^2 n^2$ for $n = 1$ to infinity and 0 otherwise. Now, let us find the mean.

The first moment, μ'_1, is nothing but the expected value of X, which is defined as the summation of $x_k P(X = x_k)$. Here, we have an infinite series, so the summation is from $n = 1$ to infinity. The values of x_k correspond to n, and $P(X = x_k)$ is equal to $P(X = n)$. This can be expressed as the expected value of $X = \Sigma (n \times P(X = n)) = \Sigma (n \times 6/\pi^2 n^2)$ from $n = 1$ to infinity. You can see that the constant $6/\pi^2$ can be factored out, resulting in $\Sigma (1/n)$.

However, this series diverges; therefore, we cannot find the mean. Since the mean does not exist, higher-order moments will also not exist. For instance, μ'_2, which involves summing n^2 times the probability, will also be divergent. Thus, the mean, μ'_1, does not exist, and consequently, the variance, σ^2, also does not exist. This example shows that for a discrete random variable, the mean does not exist, and thus the variance also does not exist.



Now, let's discuss how to find the mean and variance of a continuous random variable. We have previously discussed some examples, and now we'll take another example. Let’s consider a uniform distribution. We have already talked about the mean and variance in this context. For this case, we may have touched on it but didn't go into detail, so let’s consider a function where X is a continuous random variable with the probability density function given by:

$f(x) = 1/4$, for $-2 < x < 2$,

$f(x) = 0$, otherwise.

To find the mean of this random variable, we use the expected value of X, denoted as μ_1'. This is defined by the integral:

$$\mu_1' = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx$$

Since f(x) is non-zero only when $-2 < x < 2$, we rewrite this as:

$\mu_1' = \int_{-2}^{2} x \cdot 1/4 \, dx$

This simplifies to:

$\mu_1' = (1/4) \int_{-2}^{2} x \, dx$

The integral of x is $x^2/2$, so we evaluate this from x = -2 to x = 2:

$\mu_1' = (1/4) [(2^2)/2 - (-2^2)/2] = (1/4) \cdot (2 - 2) = 0$

Thus, the mean $\mu_1' = 0$.

Now, to find the variance σ^2, we use the formula:

$\sigma^2 = E[(X - \mu_1')^2]$

Since $\mu_1' = 0$, this simplifies to:

$\sigma^2 = E(X^2) = \mu_2'$

To find μ_2', we calculate:

$\mu_2' = \int_{-2}^{2} x^2 \cdot f(x) \, dx = \int_{-2}^{2} x^2 \cdot 1/4 \, dx$

This simplifies to:

$\mu_2' = (1/4) \int_{-2}^{2} x^2 \, dx$

The integral of x^2 is $x^3/3$, so we evaluate this from x = -2 to x = 2:

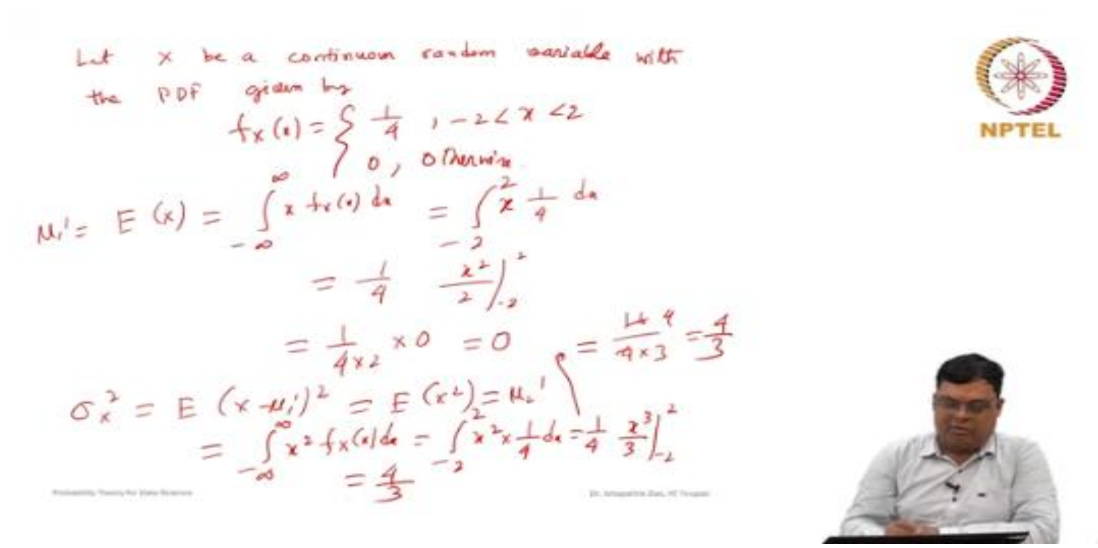
$\mu_2' = (1/4) [(2^3)/3 - (-2^3)/3]$

This gives:

$\mu_2' = (1/4) [8/3 - (-8/3)] = (1/4) \cdot (16/3) = 4/3$

Thus, the variance $\sigma^2 = 4/3$.

I hope you have understood how to find the mean and variance for discrete random variables and continuous random variables. We also provided an example of a discrete random variable where the mean and variance do not exist.



Now, let us consider this continuous random variable and its density function. Here, we have the density function that we discussed. So, X is a continuous random variable defined by $f(x) = c / x^2$. We found that $c = 1$, so the density function is $f(x) = 1 / x^2$ for $x > 1$ and $x < \infty$, and it is 0 otherwise. Let X be a continuous random variable with the probability density function given by:

$f(x) = 1 / x^2$, for $x > 1$ and $x < \infty$,

$f(x) = 0$, otherwise.

We also found the probability between 2 and 3. Now, what will be the mean of this random variable? μ_1', which is the expected value of X, is defined by the integral:

$\mu_1' = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx$

Because $f(x)$ is non-zero only from $x = 1$ to $x = \infty$, this becomes:

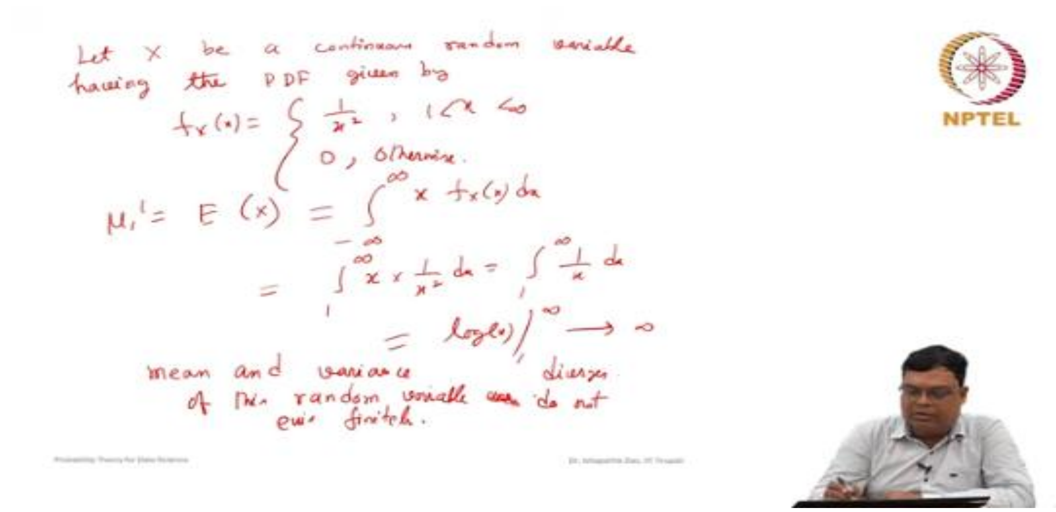
$\mu_1' = \int_{1}^{+\infty} x \cdot (1 / x^2)\, dx$

This simplifies to:

$\mu_1' = \int_{1}^{+\infty} (1 / x)\, dx$

Integrating $(1 / x)$ gives $\log(x)$. If you evaluate this from 1 to ∞, it diverges. So, the mean does not exist finitely; it is actually divergent. Since the mean does not exist, higher-order moments will also not exist here. Therefore, this is an example of a continuous random variable where the mean and variance do not exist finitely.

In summary, we can say that the mean and variance of this random variable do not exist. So, this is what we discussed in the moment. I hope it is clear now how to find the first order moment, second order moment, mean, and variances for discrete random variables and continuous random variables. We have discussed some numerical examples as well. Next, we will look at some special distribution functions that we usually use when modeling or analyzing data



When we find data in nature, we often find these distributions to be very useful. The Bernoulli distribution, binomial distribution, and Poisson distribution are examples of discrete distributions. On the continuous side, we have distributions like the exponential distribution, gamma distribution, and normal distribution. We will explicitly discuss how these distributions look, what their means and variances are, and cover their distribution functions, probability mass functions in the case of discrete random variables, and probability density functions for continuous random variables. It's important to learn these distributions well because they are frequently used in statistical inference and data science.

One of the distributions we will start with is the Bernoulli distribution. From the beginning, we discussed a special example: tossing a coin. When tossing a coin, the outcomes are either heads or tails. In this case, we generally assume it is an unbiased coin, so the probability of getting heads is 1/2. However, it may not always be exactly 1/2; we will assume that in general, the probability of heads is some p, where $p \in [0, 1]$.

In nature, we find this kind of random phenomenon as well. For instance, when we give medicine to a patient, the response can often be categorized into two outcomes: success or failure. We can model this data using the Bernoulli distribution. When tossing a coin once, the sample space consists of heads and tails. We define a random variable X such that:

$X = 1$ for heads

$X = 0$ for tails.

The probability mass function is defined accordingly. The relationship is that $X = 1$ for heads and $X = 0$ for tails. Now, let's determine the probability that $X = 1$. We assume this probability is p, which may not always be 1/2. So, in general, this is represented as p.

The probability that it results in tails is $1 - p$. Thus, the probability mass function can be expressed as follows. In general, we denote the probability mass function of this random variable as $P_X(x)$, which can also be represented as $P(x)$. This is equal to:

$$P_X(x) = p^x * (1 - p)^{(1 - x)}, \text{ where } x \in \{0, 1\}$$

Otherwise, it is equal to 0.

To clarify the relationship, consider that $P(X = 1)$, which is the probability mass function. When you substitute $x = 1$, you get:

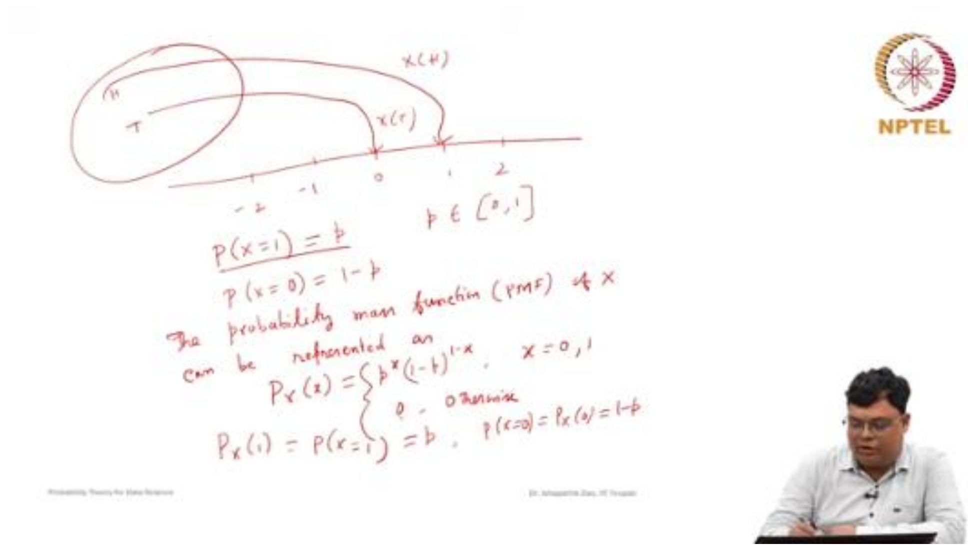
$$P_X(1) = p^1 * (1 - p)^0 = p.$$

Thus, the probability that $X = 1$ is p. Now, let's look at the probability that $X = 0$. The probability mass function for $X = 0$ is $P(X = 0)$.

If you substitute $x = 0$, this becomes:

$$P_X(0) = p^0 * (1 - p)^1 = 1 * (1 - p) = 1 - p.$$

For any other value, the probability will be zero. The range of X is only 0 and 1. Now, if you draw the probability mass function, it is very simple.

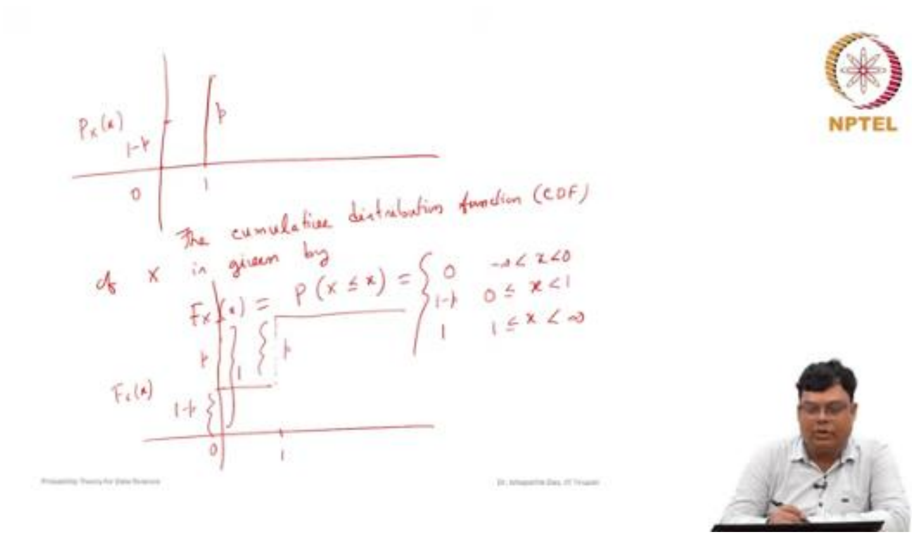


This may be useful when we discuss other random variables. The probability mass function P(X) at 0 is 1 - p, and at 1, it is p. Here, we assume that p > 1 - p. Next, what will be the cumulative distribution function (CDF)? The cumulative distribution function of X is given by F(X), which represents the probability that $X \leq x$.

To find this, we take the probability mass function, which only considers 0 and 1. For any value of $X < 0$, since the minimum value of this random variable is 0, the cumulative distribution function will be 0. Whenever $X \geq 0$, then $P(X = 0)$ is equal to 1 - p, and this applies up to 1. We have discussed previously how to find the probability distribution function from the probability mass function. Thus, whenever X is less than the minimum value, the cumulative distribution function will equal 0.

When X is between 0 and 1, the value of $P(X = 0)$ is 1 - p. For values greater than 1, the cumulative distribution function will be the maximum value, which is $P(X = 0) + P(X = 1)$. Therefore, this is $(1 - p) + p$, which equals 1. This is less than infinity because there are only two possible values. To graphically represent the cumulative distribution function, you would illustrate it as follows: for values less than 0, the CDF is 0.

At 0, it is equal to 1 - p, and this continues up to 1. At 1, it will equal 1. If you write this as 1, the distance is p. Thus, the total distance will be 1, meaning that $p + (1 - p) = 1$. This is a step function because it is a discrete random variable. It will be a step function, representing the cumulative distribution function.



Now, if we want to find the r-th order moment, as well as the mean and variances, we start by calculating the first-order moment, denoted as μ_1. The first-order moment, μ_1, is the expected value of X. By definition, this is given by $\mu_1 = \Sigma x_k P(X = x_k)$. Since this random variable takes only two values, 0 and 1, we have $\mu_1 = 0 \times P(X = 0) + 1 \times P(X = 1) = 0 + 1 \times P(X = 1) = P(X = 1) = p$. Now, we consider the second-order raw moments.

In general, to find the r-th order raw moment, we use the definition $\mu_r = E[X_r] = \Sigma x_{kr} P(X = x_k)$. For our case, where the values are only 0 and 1, we get $\mu_r = 0_r \times P(X = 0) + 1_r \times P(X = 1) = 0 + 1 \times P(X = 1) = p$. Thus, for any r from 1 to n, not just for r = 1, this value is p. This means that for r = 2, $\mu_2 = p$. It's important to find these moments because we want to calculate the variance of this random variable, which is defined as $Var(X) = E[(X - \mu_1)^2]$.

However, we also have a simplified formula: $Var(X) = \mu_2 - \mu_1^2$. Since $\mu_2 = p$, we can substitute: $Var(X) = p - \mu_1^2 = p - p^2 = p(1 - p)$. This is the variance of the Bernoulli random variable, and we have discussed how simple it is to find the mean and variances in this case. You can also compute the r-th order moment. So, this is one of the important distributions we discussed here, including their probability mass function, how it looks, and what the probability distribution function and cumulative distribution function are.

$$\mu_1' = E(X) = \sum x_k \, p_X(x_k)$$
$$= 0 \times p_X(0) + 1 \times p_X(1)$$
$$= 0 + 1 \times p = p$$

$r = 1, 2, \ldots$

$$\mu_r' = E(X^r) = \sum x_k^r \, p_X(x_k)$$
$$= 0^r \times p_X(0) + 1^r \times p_X(1)$$
$$= 0 + 1 \times p = p$$

$r = 2, \; \mu_2' = p$

$$V(X) = \sigma_X^2 = E(X - \mu_1')^2 = \mu_2' - \mu_1'^2$$
$$= p - p^2 = p(1-p)$$