Artificial Intelligence (AI) for Investments

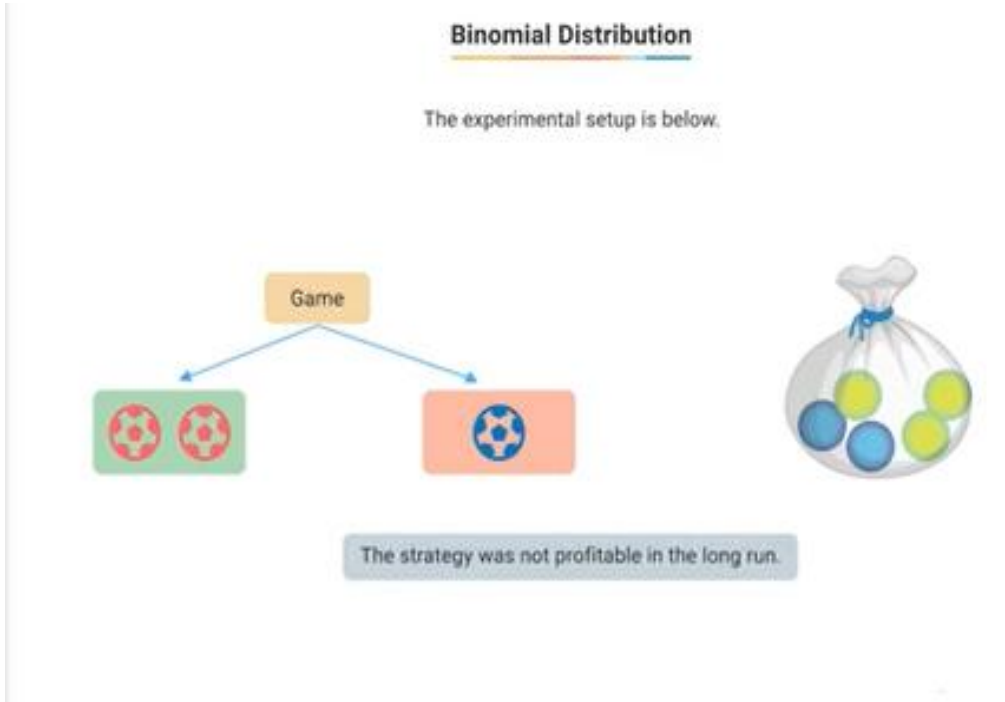
Prof. Abhinava Tripathi

Department of Industrial and Management Engineering

Indian Institute of Technology, Kanpur

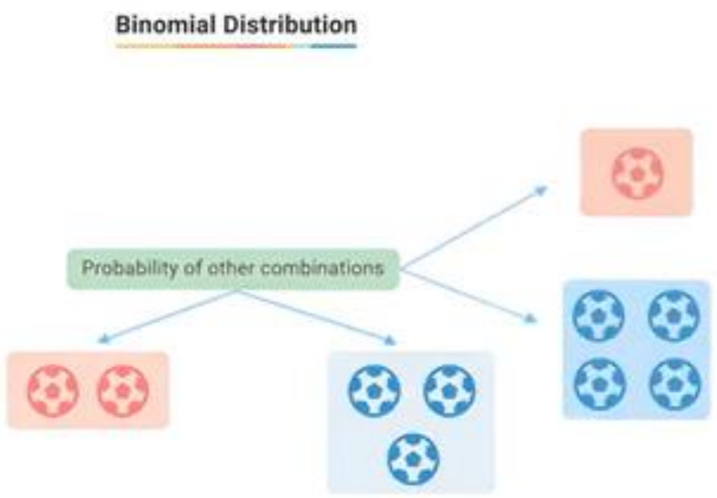
Lecture - 13

(Refer Slide Time: 00:16)



Binomial distribution part one. Let us go back to our experimental setup where we played a game with two red and one blue balls. You saw that by asking people to choose four balls from the bag we were able to evaluate the probability and conclude that our strategy would not be profitable in the long run.

(Refer Slide Time: 00:30)



But what if you want to evaluate profitability of other combinations like two red balls and three blue balls one red ball and four blue balls.

(Video Starts: 00:39)

Is there a way we can generalize these experiments? Let us go through a very commonly occurring and useful probability distribution the binomial distribution using the same red and blue ball example we will see how such problems can be generalized and solve using binomial distributions. Let us start with some simple questions. Suppose you don't have the resources to conduct the experiment where people draw four balls from a bag containing two red and one blue ball.

Can you still find out the probability of getting one two or three red balls? Let us try it out. Now, since there are two red balls and one blue ball in the bag the probability of getting a red ball in one trial is $\frac{2}{3}$ and the probability of getting a blue ball is $\frac{1}{3}$. Suppose you are drawing four balls from the bag. Can you tell us the probability of getting a red ball each time?

Remember that the multiplication rule states that the probability of event 1 and event 2 happening if they are independent is equal to probability of event 1 multiplied by probability of event 2. Here we have four events taking place as we are drawing four red balls hence the probability of all these four events taking place is going to probability of event 1 multiplied by probability of event 2 multiplied by probability of event 3 and multiplied by probability of event 4.

According to the multiplication rule of probability the probability of getting four red balls after four trials equal to probability of getting a red ball in the first trial multiplied by probability of getting a red ball in the second trial multiplied by probability of getting a red ball in the third time multiplied by probability of getting a red ball in the fourth trial which is equal to $\frac{2}{3} * \frac{2}{3} * \frac{2}{3} * \frac{2}{3} = 0.197$. Remember that we are replacing the red ball again after drawing it.

Now suppose we do not the event of drawing a red ball as upper case R and the event of drawing a blue ball as upper case B. Then the combination of getting all four red balls will be four times R. Now let us consider another possible scenario where the sequence of drawing the balls is B times B and triple R. This means you drew a blue ball in the first trial and red balls in the remaining three trials, what would be the probability in this case?

Once again applying the multiplication rule to calculate the probability we would get the probability as probability of drawing a blue ball which would be $\frac{1}{3}$ times probability of trying the red ball which could be $\frac{2}{3}$ times. Probability of trying a red ball which would be again $\frac{2}{3}$ times probability of trying a red ball. And this would be equal to $\frac{1}{3}$ multiplied by $\frac{2}{3}$ multiplied by $\frac{2}{3}$ multiplied by $\frac{2}{3}$ equal to 0.0987.

Remember that $\frac{1}{3}$ is the probability of getting a blue ball in one trial and $\frac{2}{3}$ is the probability of getting a red ball in one trial. Similarly, there are various combinations in which we can get three red balls and one blue ball. Here the first ball was blue followed by three red balls but we can also have a combination where the red ball is blue and the first third and fourth balls are red and so on.

If we write down the different sequences in which we get three red balls and one blue ball then there are four such sequences that is B triple R, RB double R, double RBR and triple RB. If you calculate the probability for each of these events you will see that all the four sequences have the same probability of occurring that is 0.0987. So, the probability of getting a combination which has three red balls and one blue ball is equal to 0.0987 plus 0.0987 plus 0.0987 that is four times 0.0987 which is equal to 0.3948.

Notice that we are using the additional rule here, the addition rule states that the probability of two events happening that is event 1 or even 2 happening if they are independent is equal to probability of event 1 plus probability of event 2. Here we use the addition rule for four independent events and that is how we got 4 times 0.0987 equal to 0.3948. Now we will move on to the second part of binomial distribution.

As you may recall we can find the probability of x equal to 1. There are four possible combinations for x equal to 1. Namely, the first ball being red and the remaining three being blue or the second ball being red and the remaining three being blue and so on. Let us consider the case where the first ball is red and the other three are blue. We can compute the probability of this outcome as $\frac{2}{3}$ as the probability for the red ball times.

$\frac{1}{3}$ as the probability for the first blue ball times $\frac{1}{3}$ again as the probability for the second blue ball multiplied by $\frac{1}{3}$, that is the probability for the third blue ball. The product of all these will give us the probability of this outcome of the first red ball followed by three blue balls. The product is the value 0.0247. Since there are four such outcomes with one red ball essentially giving us the outcome of one red ball as four times 0.247 which comes to 0.0988.

Similarly, we can compute the probabilities of getting 0, 2, or 4 red balls in the same way. Let us try to generalize this probability let us say that the probability of getting a red ball in one trial is not $\frac{2}{3}$ but P what would be the probabilities in that case? Again, P is the probability of getting a red ball the probability of getting four bread balls is in four trials equal to $(p)^4$.

Similarly, since there are four combinations in which we can get three red balls the probability for x = 3 is $4 * (p^3 * (1 - p))$. Why did we get this 1 - p? 1 - p represents the probability that we are getting a blue ball. So, we evaluate this expression as $4p^3 * (1 - p)$. The probability that the random variable will take the value 0 is given as $(1 - p)^4$.

Why so? This is because 0, red ball means there are four blue balls and each blue ball has a probability of 1 - p, thus it is 1 - p multiplied by 1 - p multiplied by 1 - p and so on up till $(1 - p)^4$.

(Video Ends: 06:22)

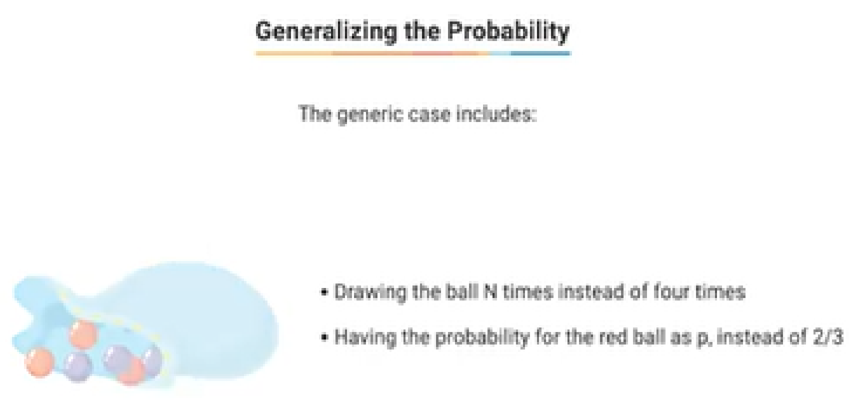
(Refer Slide Time: 06:22)

Generalizing the Probability

Combination	Probability
P(X=0)	$(1-p)^4$
P(X=1)	4*p(1-p)
P(X=2)	$6p^2(1-p)^2$
P(X=3)	$4p(1-p)^3$
P(X=4)	p^4

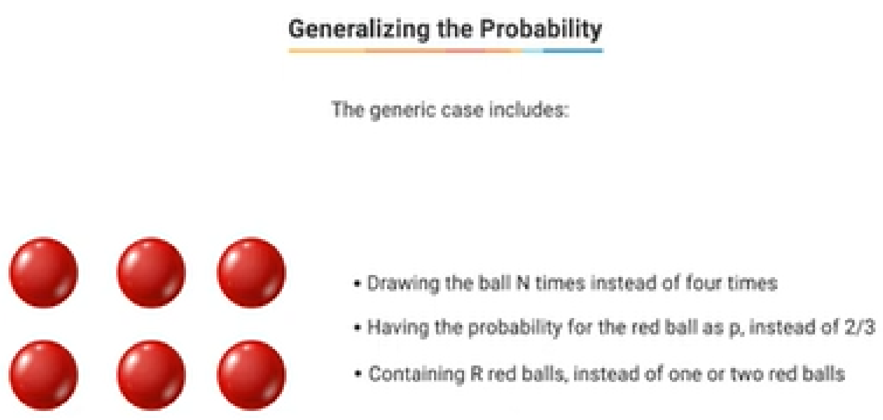
Probability of taking the value 1 for r random variable is $4 * p\ (1 - p)$. Again, this is because there are four combinations where we can have one red ball and three blue balls. The probability of X = 2 is $6p^2(1 - p)^2$ 6 p to the power 2 multiplied by 1 - p to the power 2 for X = 3 we had seen this earlier as $4p^3 * (1 - p)$ again for X = 4 it is just $(p)^4$.

(Refer Slide Time: 06:47)



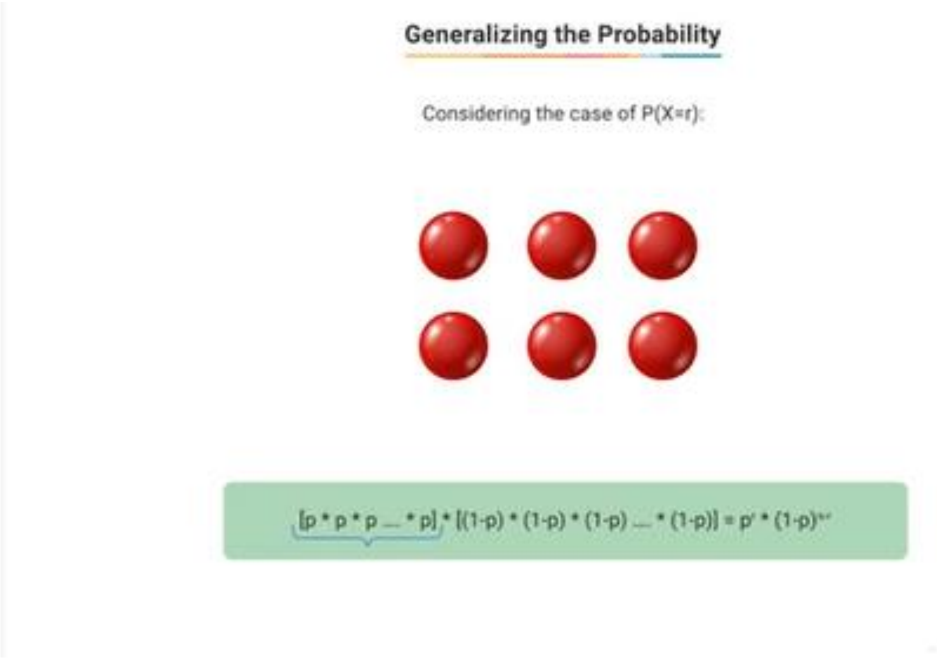
So, now let us extend this case to a more generic case where instead of four times we were to draw the ball N times instead of having probability as $\frac{2}{3}$ for the red ball let us keep it generic at a value of p and instead of having one red ball or two red balls.

(Refer Slide Time: 06:49)



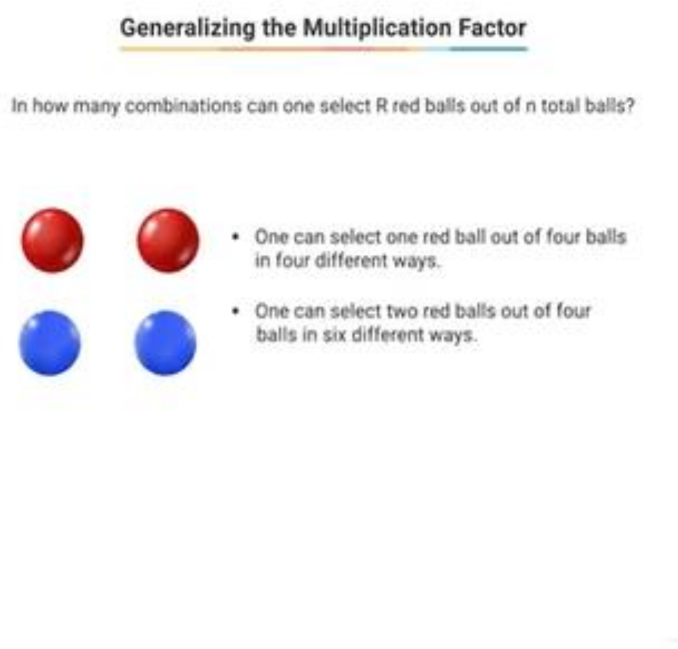
Let us generically say that we want to find the probability of getting r red balls from the back.

(Refer Slide Time: 07:06)



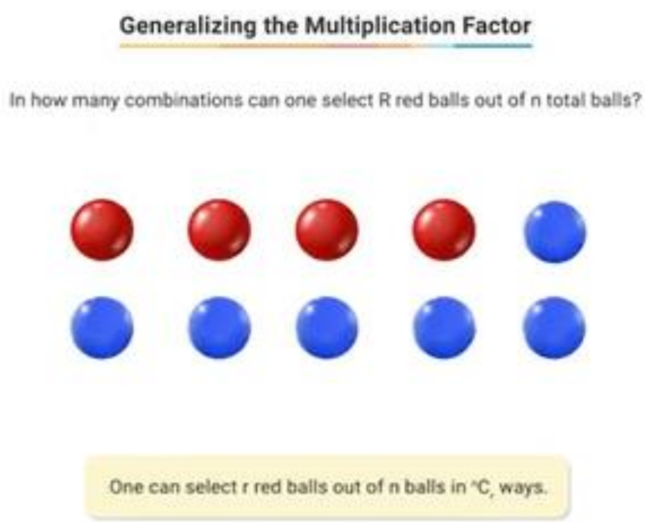
Let us see how we can calculate for P X = r, that is probability that X = r. We would have to get r red balls and N - r blue balls. So, the probability of getting one such combination is equal to $\boldsymbol{p} * \boldsymbol{p} * \boldsymbol{p} \ldots \ldots * \boldsymbol{p}$ up to r that is $\boldsymbol{p}^{\boldsymbol{r}}$ multiplied by $(\mathbf{1} - \boldsymbol{p}) * (\mathbf{1} - \boldsymbol{p}) \ldots \ldots * (\mathbf{1} - \boldsymbol{p})$ up to 1 - r times which is equal to $\boldsymbol{p}^{\boldsymbol{r}} * (\mathbf{1} - \boldsymbol{p})^{\boldsymbol{n}-\boldsymbol{r}}$. Finally let us see how we can generalize the multiplication factor.

(Refer Slide Time: 07:36)



You recall this factor was calculating the number of combinations of getting R red balls out of the n total balls. For example, when we selected one red ball out of four balls, we say that there were four different combinations in which this was possible. When we were drawing two red balls out of four balls, we saw that there were six different combinations that would generate.

(Refer Slide Time: 08:01)



So, if we have to generalize this, we can count the number of combinations in which we can get r red balls out of a total number of n balls and this will be equal to C n C r. Hence the probability of getting r red balls after n trials is equal to $nC_r p^r (1-p)^{n-r}$.

(Refer Slide Time: 08:16)

Generalizing the Multiplication Factor

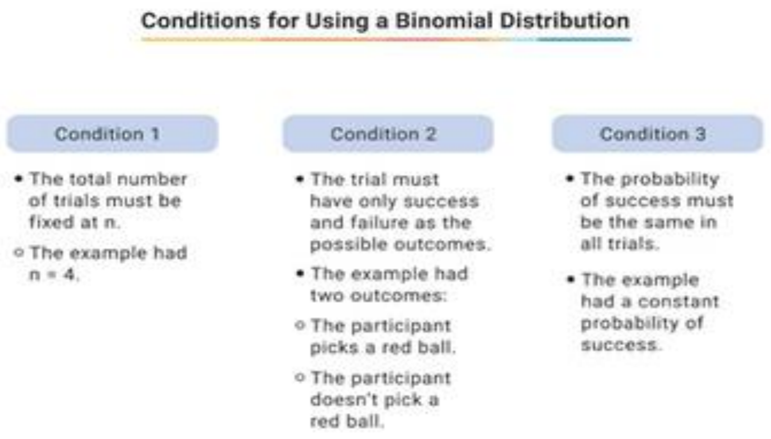
The probability distribution for different values of X can be found using the different values of R.

Combination	Probability
P(X=0)	$^nC_0\, p^0\, (1-p)^{n}$
P(X=1)	$^nC_1\, p^1\, (1-p)^{(n-1)}$
P(X=r)	$^nC_r\, p^r\, (1-p)^{(n-r)}$

The binomial probability distribution involves the mentioned formulae.

Using this formula for each value of r that is 0,1, 2, 3 till n. We can find the probability distribution of our random variable X. For example, for P X = 0 would be equal to $nC_0p^0(1 - p)^n$. P X = 1 would be equal to $nC_1p^1(1-p)^{n-1}$ and so on up till P X = r would be equal to $nC_rp^r(1 - p)^{n-r}$. This probability distribution is called the binomial probability distribution.

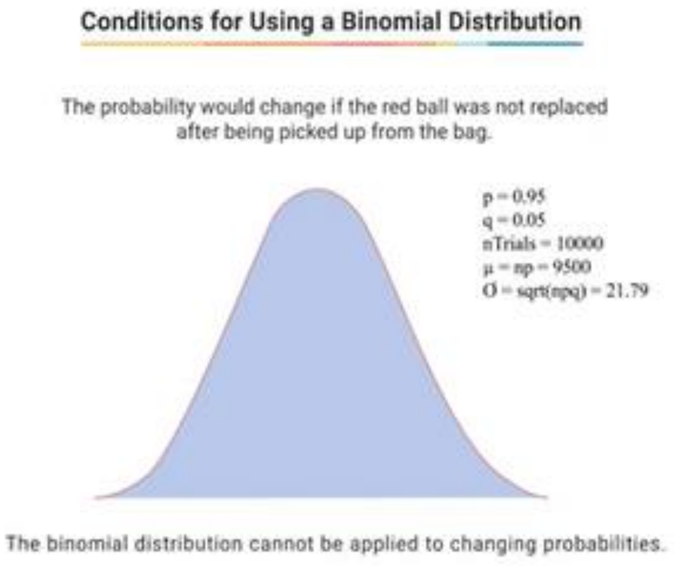
(Refer Slide Time: 08:50)



Let us now understand under what conditions we can use the binomial distribution. First the total number of trials must be fixed at n. In our example every participant got exactly four chances to pick out balls from the back so n = 4 here. The second condition is that each trial is binary nature which means that it has only two possible outcomes success or failure. Going back to our example there were only two possible outcomes that could take place when the participant tried to take out a ball.

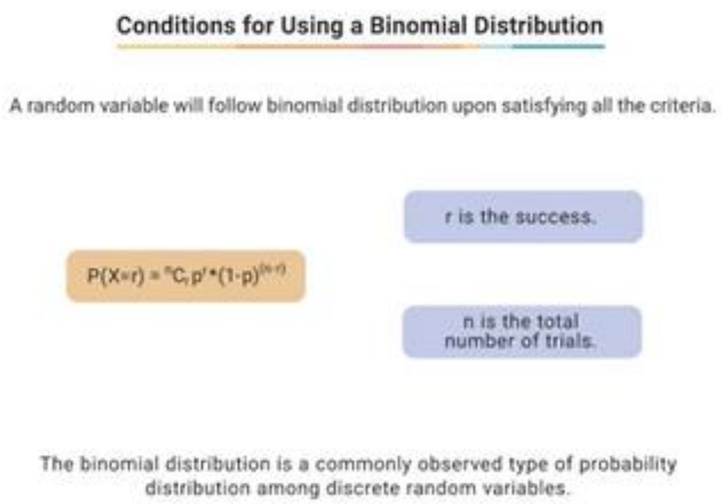
The first possibility is that he picks a red ball and the second possibility is that he does not pick a red ball. The third and final condition is that the probability of success is the same in all trials denoted by P in our case the probability of success when picking a red ball is always constant and equal to $\frac{2}{3}$.

(Refer Slide Time: 09:35)



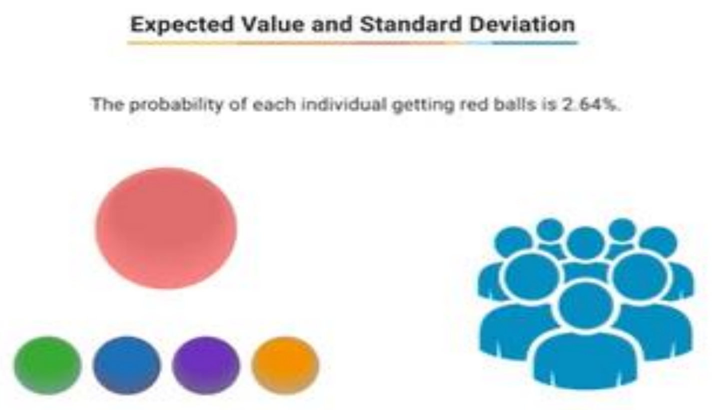
Suppose we had taken the same scenario and had not been replacing the red ball after picking it out. The first time then the probability would change then the part spend would try to pick the second red ball. In that case we would not be able to apply the binomial distribution.

(Refer Slide Time: 09:49)



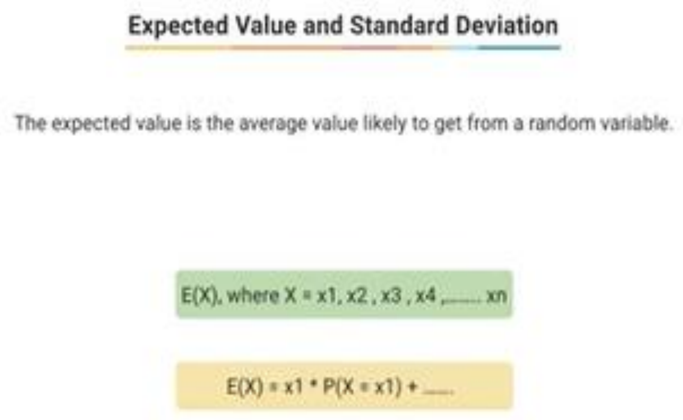
When all of these conditions are satisfied then the random variable will follow a binomial distribution and the probability for X = r that is getting r successes in n trials can be calculated as $nC_r p^r (1 - p)^{n-r}$ as given by the binomial distribution. This is a very commonly observed type of probability distribution among discrete random variables it is seen in conditions with binary results.

(Refer Slide Time: 10:20)



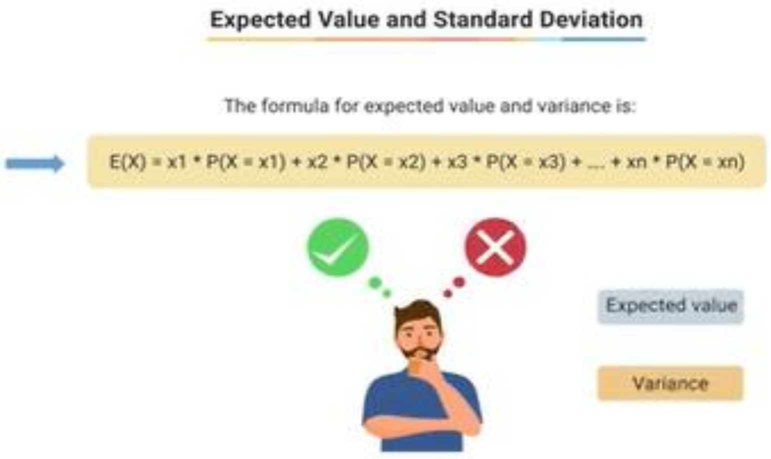
Binomial probability distribution expected value and standard deviation. Remember that in our early experiment which we conducted with 100 participants we estimated that we can expect a person to draw 2.64 red balls in one game. This value 2.64 is nothing but the expected value of the number of red balls for each individual.

(Refer Slide Time: 10:36)



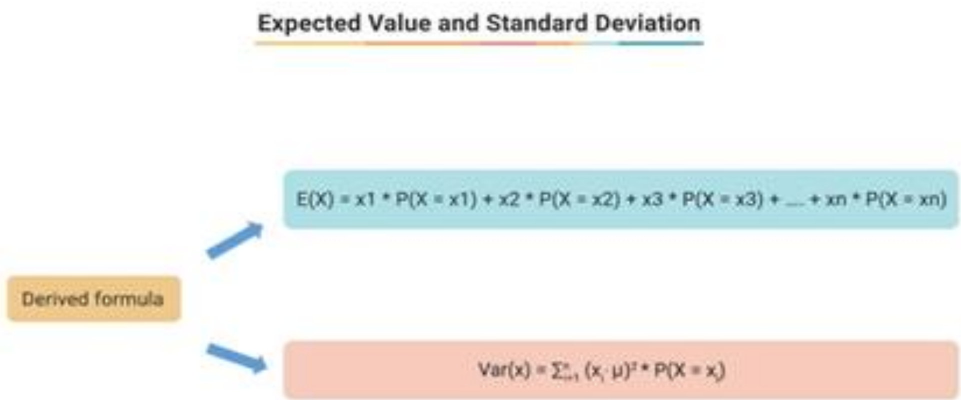
Recall that the expected value is nothing but the average value that we would expect to get from a random variable and the weight is calculated is that we take the value that X takes x 1 multiplied by the probability of that X takes x 1 summed up for all the values of X that we can take.

(Refer Slide Time: 10:53)



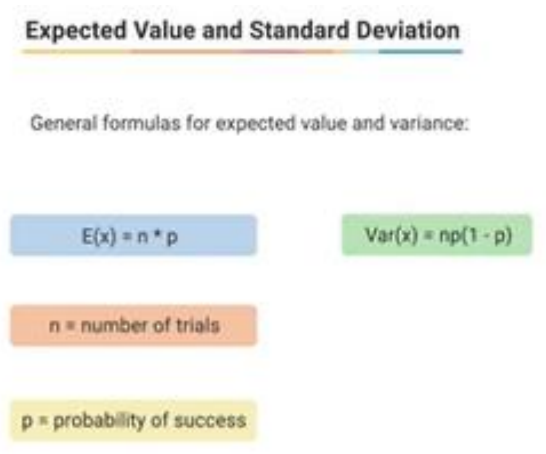
So, basically it is x 1 times probability of x 1 + x 2 times probability of x 2 and so on up till x n times probability of x n can we generalize this calculation of expected value and variance for the binomial distribution.

(Refer Slide Time: 11:06)



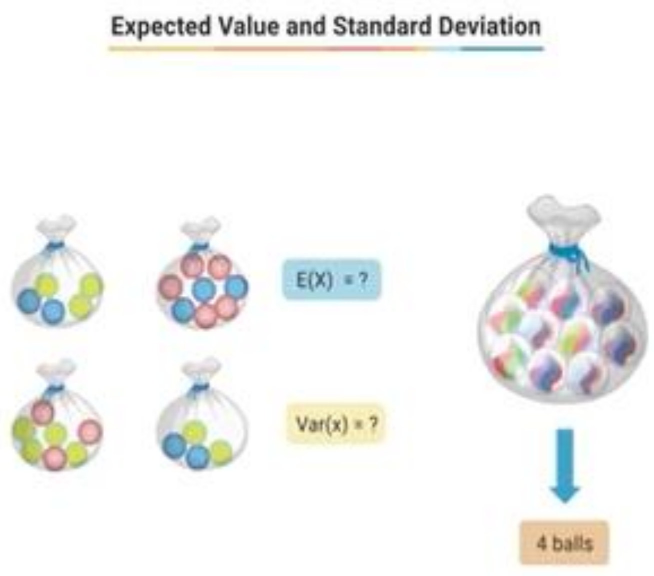
We can while the derivation of this form lies beyond the scope of the discussion for any random variable that is known to follow a binomial distribution.

(Refer Slide Time: 11:14)



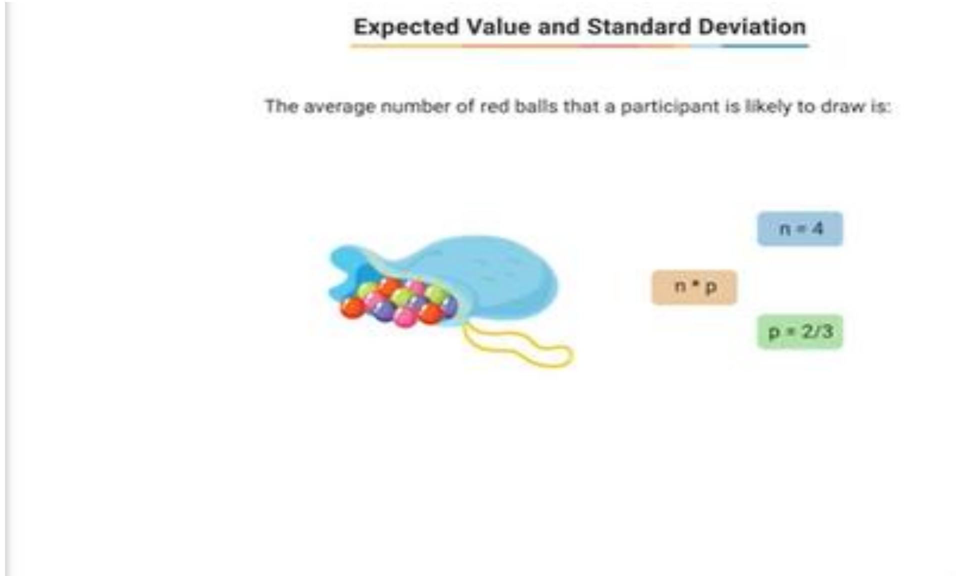
We can compute the expected value and variance using these general formulas. The expected value of x will be n times p where n is the number of trials and p is the probability of success then we have variance which is equal to $np(1-p)$.

(Refer Slide Time: 11:31)



So, let us take our example of calculating the number of red balls and calculate the expected value and variance. Remember that in any experiment the participant was allowed to draw four balls.

(Refer Slide Time: 11:41)

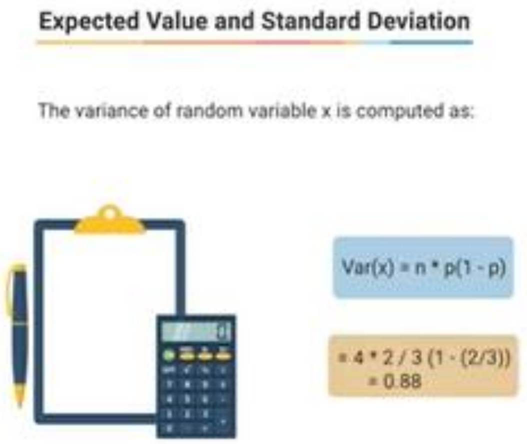


So, how many red balls on average can we expect participant to draw? The answer will be n time p where n = 4 and p = $\frac{2}{3}$ as that is the probability of getting a red ball.

(Refer Slide Time: 11:50)

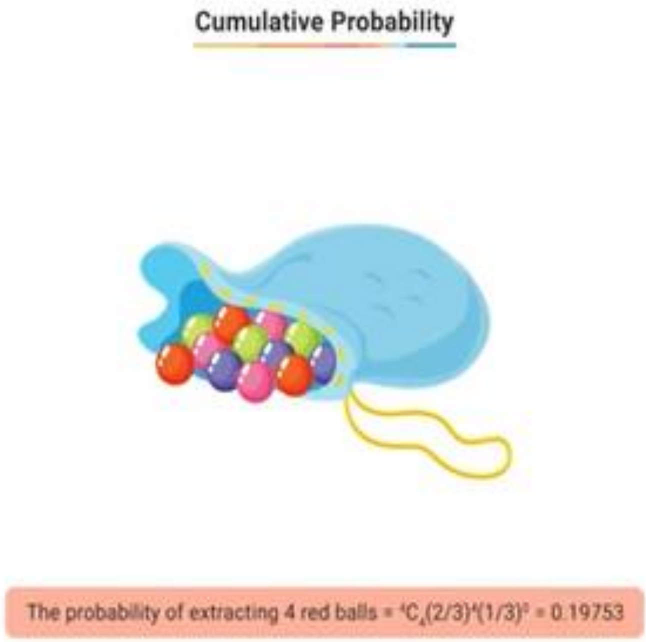
Multiplying both these values we get 4 times $\frac{2}{3}$ which is 2.67 this means on average we can expect the participant to draw 2.67 red balls in a game.

(Refer Slide Time: 12:02)



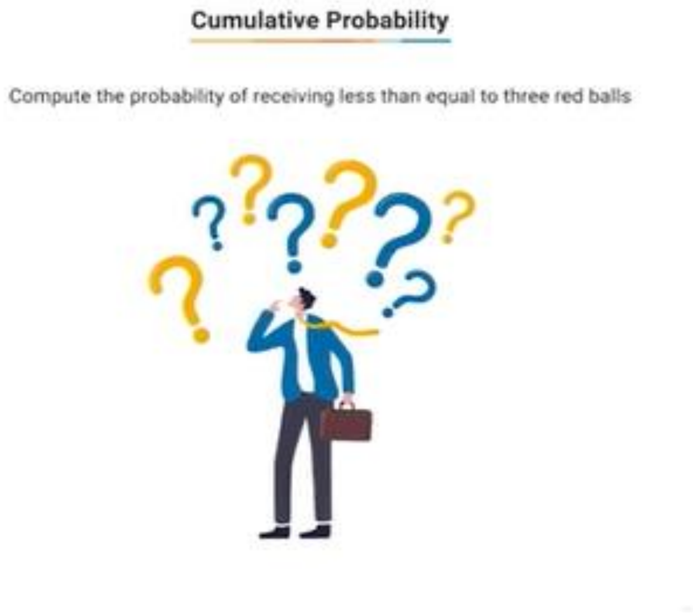
Similarly, we can calculate the variance of this random variable x this would be an $np(1 - p)$, that is going to be $4 * \frac{2}{3} * \frac{1}{3} = 0.88$.

(Refer Slide Time: 12:15)



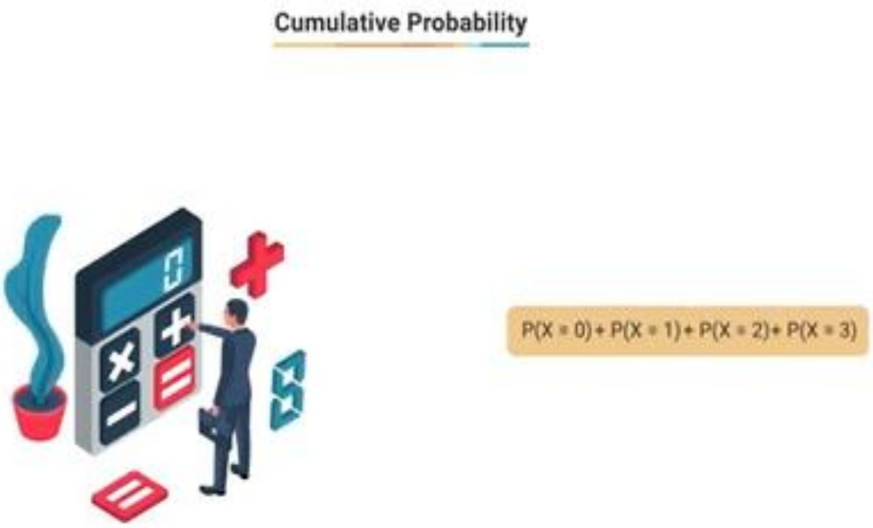
You saw that using the binomial distribution. We were able to calculate the probability of getting an exact value. For example, if you want to know the probability of X = 4 that is 4 red balls. We know the formula that is $4C_4\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^0 = 0.1975$.

(Refer Slide Time: 12:36)



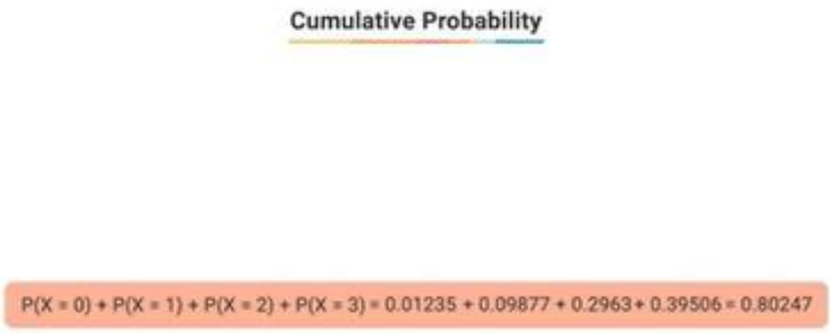
But what would you do if you wanted to calculate the probability of getting less than equal to 3 red balls.

(Refer Slide Time: 12:42)



Therefore, we are calculating the probability that X can take value 0,1,2,3. We can say that this value is simply calculated as probability of X = 0 plus probability of X = 1 plus probability of X = 2 plus probability of X = 3.

(Refer Slide Time: 12:57)

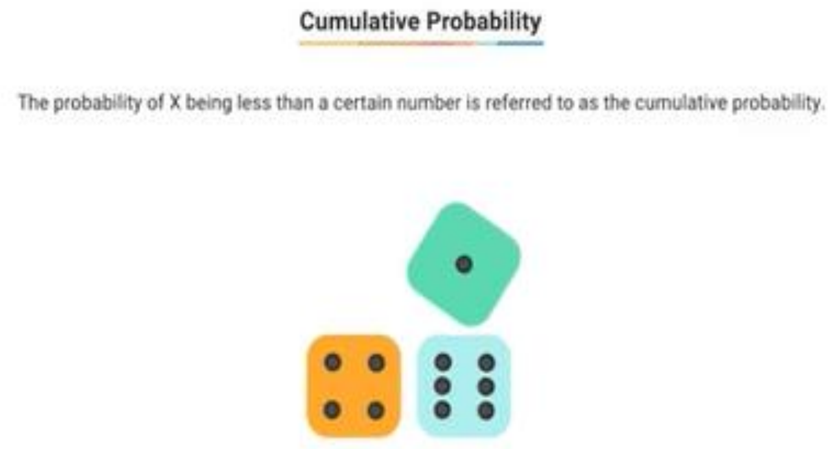


We can substitute these values and we will find that is equal to 0.1235 plus 0.9877 plus 0.2963 plus 0.395 which is equal to 0.8024.

(Refer Slide Time: 13:11)

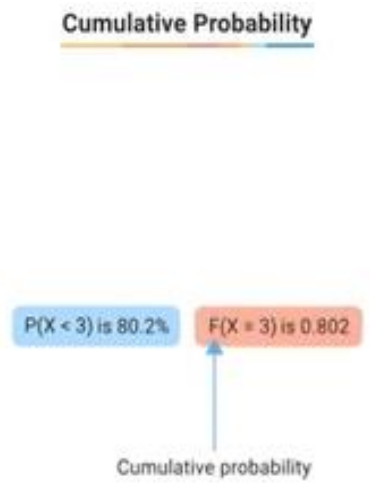
Thus, there is an 80.2 percent chance that any randomly selected participant will have selected at max 3 red balls while drawing four balls.

(Refer Slide Time: 13:19)



These kinds of probabilities are very frequently calculated in real life. So, any probability where we have to determine the likelihood of X being less than a certain number is called cumulative probability.

(Refer Slide Time: 13:29)



For example, 0.802 is the cumulative probability for X less than equal to 3 and instead of saying probability P of X Less than 3 is 80.2 percent we can use $F(X = 3)$ that is 0.802 where F represents the cumulative probability.

(Refer Slide Time: 13:47)

Cumulative Probability

X	P(X = x)	F(X = x)
0	$= 4C_0 \cdot \left(\frac{2}{3}\right)^0 \cdot \left(\frac{1}{3}\right)^4 = 0.01235$	0.01
1	$= 4C_1 \cdot \left(\frac{2}{3}\right)^1 \cdot \left(\frac{1}{3}\right)^3 = 0.09877$	0.11
2	$= 4C_2 \cdot \left(\frac{2}{3}\right)^2 \cdot \left(\frac{1}{3}\right)^2 = 0.2963$	0.41
3	$= 4C_3 \cdot \left(\frac{2}{3}\right)^3 \cdot \left(\frac{1}{3}\right)^1 = 0.3951$	0.80
4	$= 4C_4 \cdot \left(\frac{2}{3}\right)^4 \cdot \left(\frac{1}{3}\right)^0 = 0.1975$	1.00

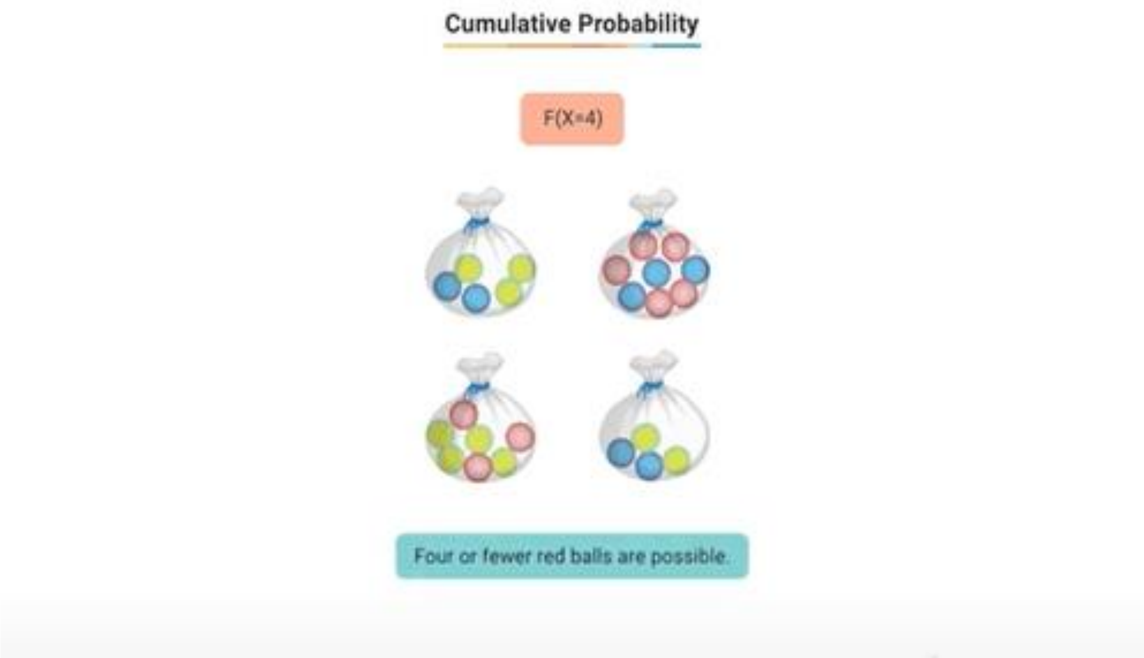
Similarly, we can calculate the cumulative probability for each value of X and create a table. And shown here you can see the probability distribution table, that we had originally created using this we can see how to create cumulative probability table for X = 0 the cumulative probability will be probability of X less than equal to 0 since for X less than equal to 0, X can take only one value which is 0. The cumulative probability here will be same as probability of X = 0 which is 0.01235.

Next let us calculate $F(X = 1)$ where $F(X = 1)$ means probability of X less than equal to 1 which will be $P(X = 0) + P(X = 1) = 0.11$. Similarly, we have $F(X = 2)$ which will be $P(X = 0) + P(X = 1) + P(X = 2)$. Now we already know that $P(X = 0) + P(X = 1)$ is as this is the same as $F(X = 1)$.

Thus, we can write $F(X = 2)$ as $F(X = 1)$ + $P(X = 2)$. We know that $F(X = 1)$ is 0.11 and we know that $P(X = 2)$ is 0.296 hence $F(X = 2)$ will be $0.111 + 0.2963 = 0.4074$.

Similarly, $F(X = 3)$ will be $0.4074 + 0.39506 = 0.80247$ or 80.247 percent and $F(X = 4)$ will be $F(X = 3) + P(X = 4)$ which is $0.80247 + 0.1975 = 1$.

(Refer Slide Time: 15:24)



This is because $F(X = 4)$ means that we are finding the probability that out of four balls drawn four or less could be red balls.

(Refer Slide Time: 15:30)

This will always be true as the number of red balls can never be more than four.

(Refer Slide Time: 15:34)

If the maximum boss that we are drawing is 4. Hence the final value in the cumulative distribution table will always be 1.