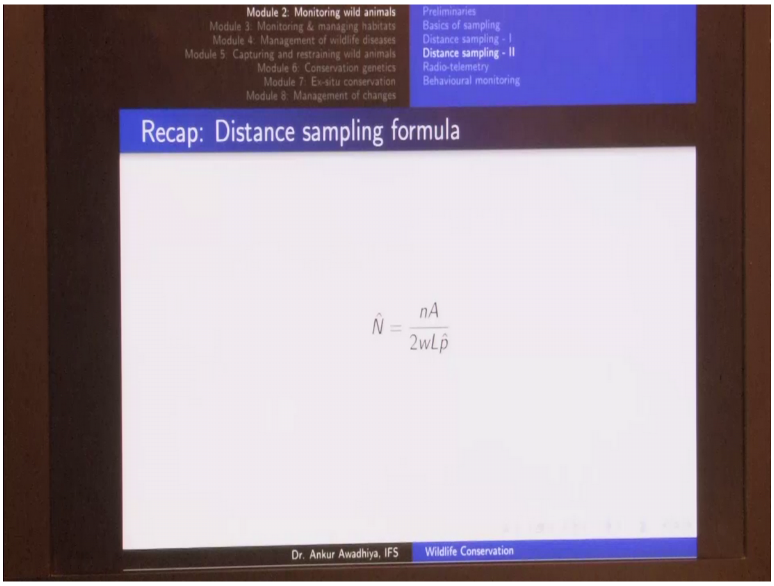
Wildlife Conservation
Dr. Ankur Awadhiya
Indian Forest Services, M.P
Indian Institute of Technology, Kanpur

Lecture – 8
Distance Sampling – II

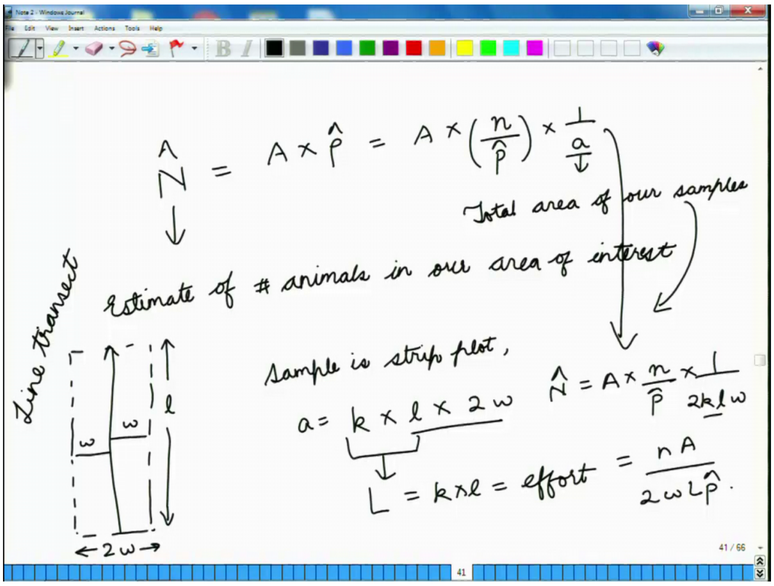
[FL] In today's lecture we continue our discussion on Distance Sampling.

(Refer Slide Time: 00:23)



So, as we saw in our previous lecture this is the distance sampling formula as you can see on your screens, N hat is equal to n A by 2 w L p hat.

(Refer Slide Time: 00:37)



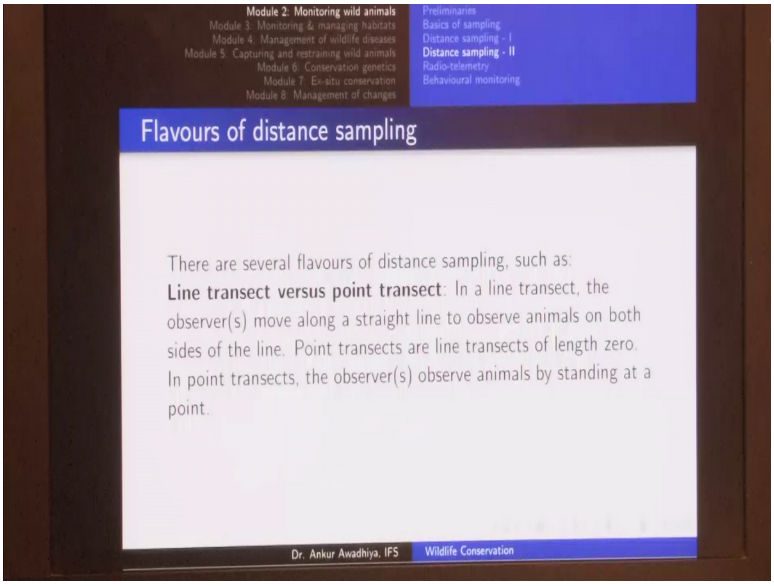
So, what essentially it is telling us is that N hat N hat which is in estimate of number of animals in our area of interest. And this figure N hat, hat because it is in estimate, it is given by area into an estimate of the density.

Now, area into the estimate of the density is given by the number of animals that we saw there as modified by the detection probability. So, it is n by p hat into 1 by a and a is the total area of our samples. Now in case the sample is strip plot we have a is given by the number of samples into the length into twice the width.

So, what we are referring to here is that, here is our transect line and this is the half widths w. So, we have half width w on the other side as well and we want to calculate the area of this rectangle. So, this rectangle has a width of twice w and a length of l so, the area of the rectangle is given by l into 2 w. So, putting it back here and k into l is also written as capital L. So, capital L is k into l, which is also the effort so, effort is the total distance that we have moved.

So, putting it all back there we get N hat is equal to A into n by p hat into 1 by 2 k l w. So, we can also write it as n A by 2 k into l is given by capital L. So, you have 2 w capital L p hat, which is our formula for the distance sampling, so, for the estimate of the number of animals in our park or the area of interest. So, now distance sampling can be done in different flavors.

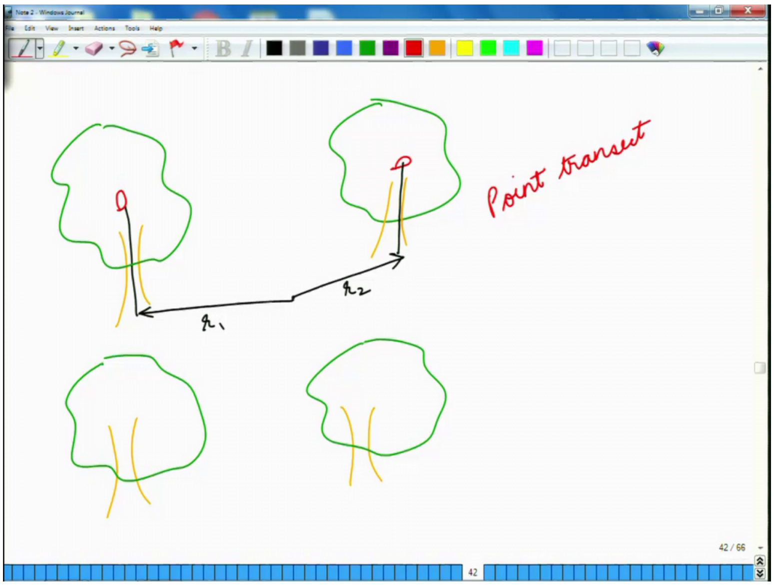
(Refer Slide Time: 03:21)



So, as you can see on your screens now so, there are different flavors of distance sampling, one is that whether do you want a line transect or a point transect. Now a line transect; is what we discussed just now in which we have moved a certain distance in a straight line so, this is our line transect.

Now, in place of a line transit; the other variation could be that we just stand at a point inside the forest. And then a a point transect rather it is generally used for birds, so, if you are standing at a point in the forest and then you have a number of trees around.

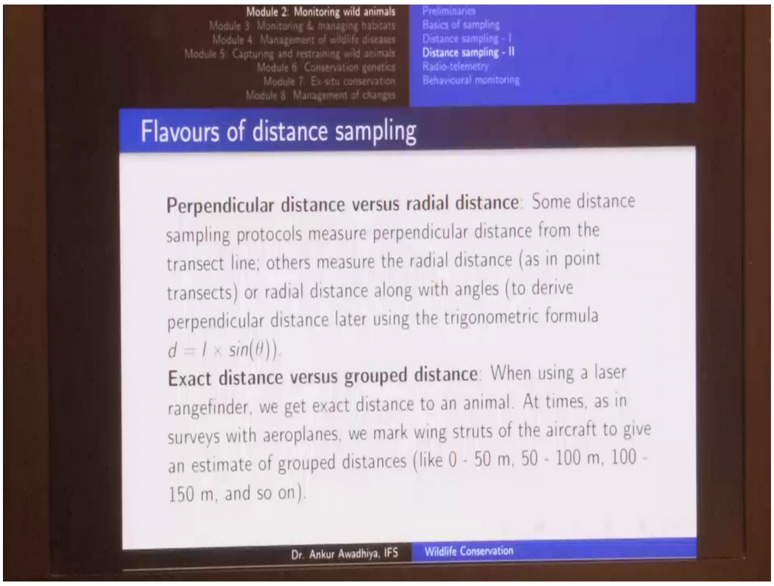
(Refer Slide Time: 03:55)



So, if there is any bird on a tree, then we are looking for the distance of the bird on the ground level. So, these are the distances as given by r 1 r 2 and so on.

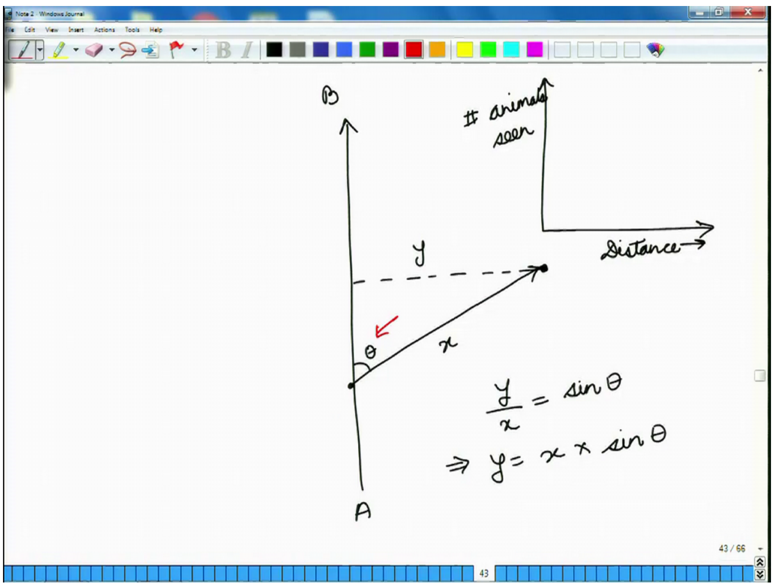
So, we can use these distances, in the similar manner as we had used these w's. So, in place of w we can have r when we have a point transect and rest of the calculations would be very similar.

(Refer Slide Time: 04:53)



Another flavor as you can see on the slides now, is perpendicular distance versus the radial distance.

(Refer Slide Time: 05:03)



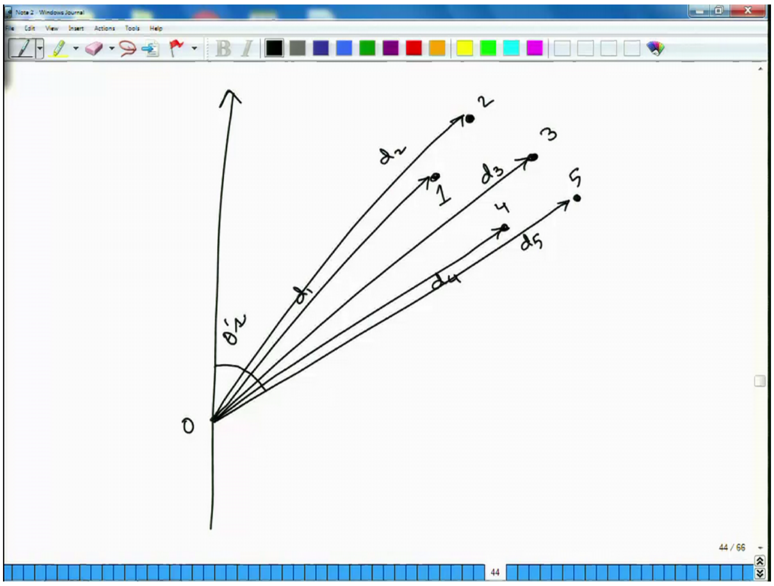
Now what do we mean here is that when we are having a line transect, so, this is the line on which we moved from A to B. Now at every point when we see an animal somewhere, we want to know it is distance from the line transect.

So, if you remember we had drawn a curve between distance from the line and the number of animals seen. Now this distance from the line, can be given by this straight line distance from the point of observation. And the angle that the animal is subtending with that the transect line or it can be given as a perpendicular distance.

Now, we know from our trigonometry, that if this distances let us call it x and if this portion is y. So, in that case, we have y by x is equal to sine theta or y can be given as x into sine theta. Where theta is the angle that is subtended between by the animal and the between the direction of the line transect.

So, when we are putting our values in the computer, we can go either for the perpendicular distance directly or we can give it in the form of a radial distance. So, you need r and you need sine theta in this case. Next we have exact distance versus group distance; what this means is that suppose we saw a group of animals.

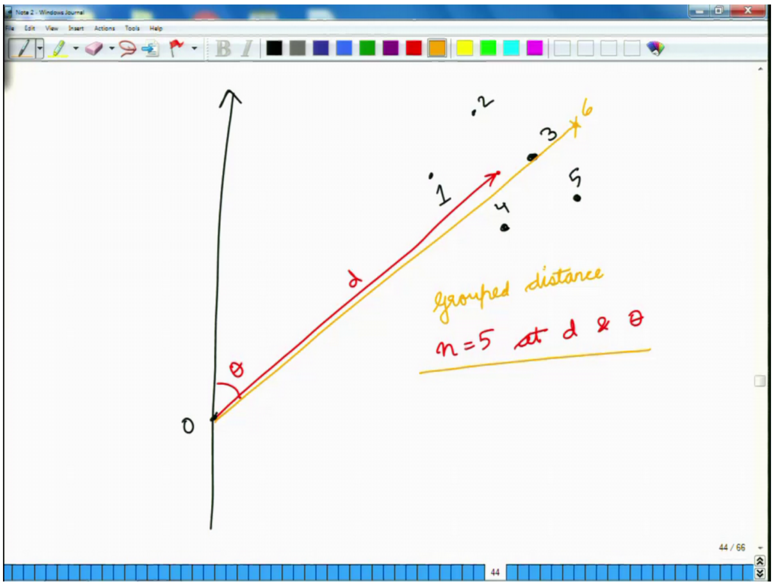
(Refer Slide Time: 06:45)



So, we have 5 animals and we are moving on this line transect and this is our point of observation o. So, when we are putting these distance values in the software, we can either go for the distance d 1 which is for the first animal.

Then a distance d 2 for the second animal, then a distance d 3 for the third animal distance, d 4 for the fourth animal and a distance d 5 for the fifth animal and all their corresponding thetas. So, now this is one way of putting our distances in the software, otherwise, we could go for one other thing we in place of putting all these different d 1 d 2 d 3 d 4 and d 5. We can just assume the center of all of these animals so let us presume that it is this point.

(Refer Slide Time: 08:02)

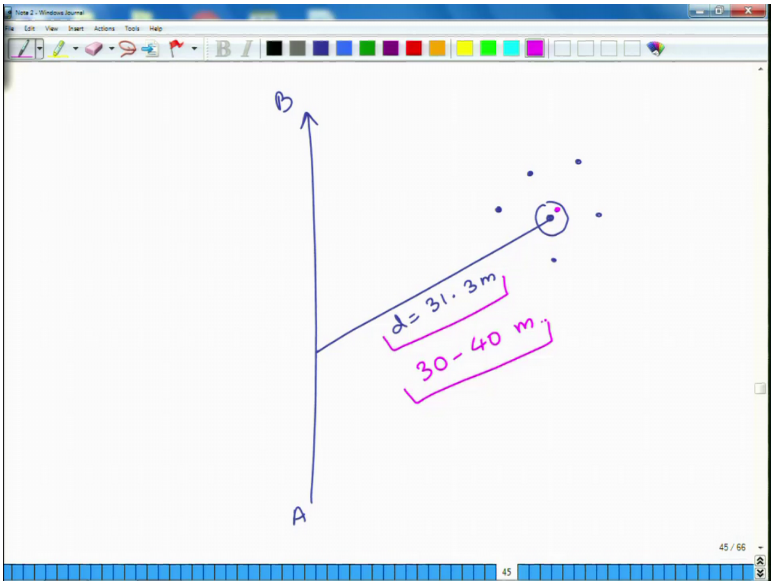


So, we can just put in a single distance d and a single angle theta and then we can put in our software that there were 5 animals n is equal to 5 at distance d and angle theta.

So, when we say exact distances versus the group distance this is what we mean, exact distances of each and every animal or a group distance for the group. Now a group distance also becomes important because, in a number of cases there might be another animal here say the sixth animal. Which we could have missed out just because, it is standing on the same line of sight as the animal number 3.

So, the animal number 6 would get occluded by the third animal so, we could miss this animal. So, it is always preferable to put distances as a group distance so, as to increase the level of precision in our calculations.

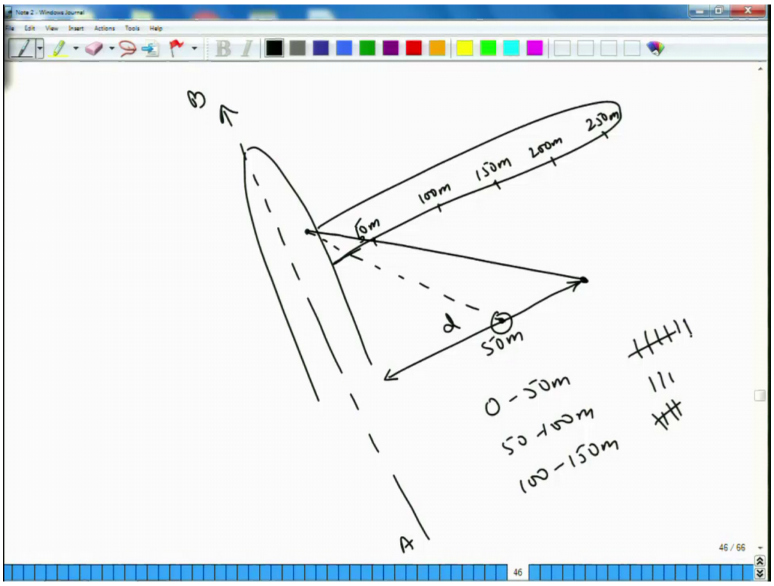
(Refer Slide Time: 09:13)



Another way in which we use the concept of group distances or clusters is when we have our transect line from A to B and when we are writing the distance of an animal or a group of animals. We can either write it as say d is equal to 31.3 meters, but in that case, we are not completely sure whether this point is the exact center your exact center could be say this point.

So, in some cases in place of writing 31.3 meters we could also group this distance as 30-40 meters. Now, in the case of distance sampling while it is always preferable to have as exact a distance as possible, in some situations group distances a become important especially in situations when we are doing an aerial survey.

(Refer Slide Time: 10:08)

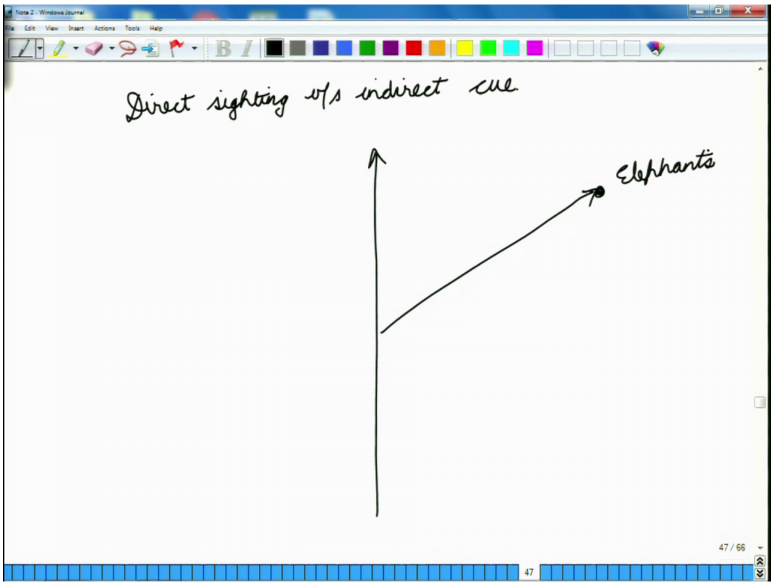


Now, in an aerial survey you have any aircraft and on the wings we have put some struts. So, the struts are basically distance estimators. So, we can just say that this is say 50 meters, this is 100 meters, this is 150 meters, this is 200 meters and this is 250 meters.

So, when we are observing a point at this particular height from the point of observation. And so, this is our transect line A to B, we saw an animal here. So, the distance between the animal and the line of the transect d would be given by this figure of 50 meters. So, this is around 50 meters from the transect line. If an animal is found here for instance so, we have only this reading that it is less than 50 meters, but we do not know exactly. Where this animal is lying because our scale that is used on the aircraft wing is not that much precise.

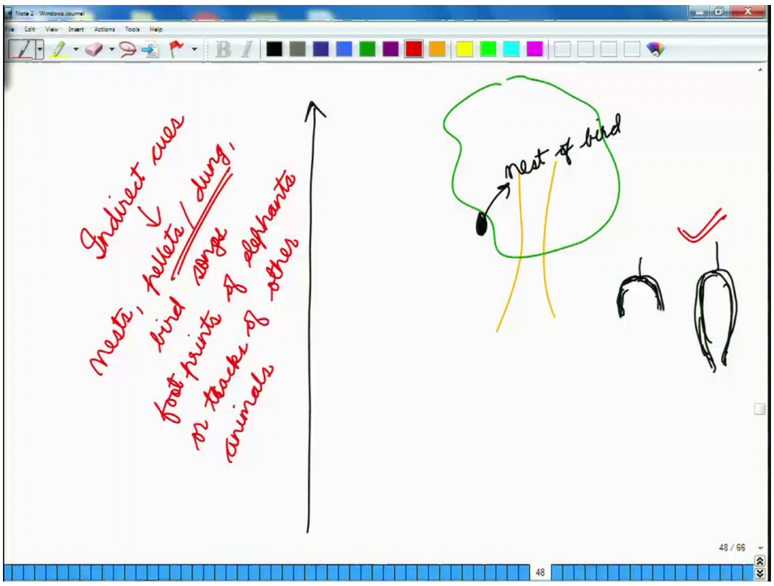
So, in that case our readings would be 0-50 meters and we saw an animal then 50-100 meters 100-150 meters and so on. So, we will just go one putting the tally marks as we see the animals. And this is another variant of putting a group distance another variation is whether we have a direct sighting or an indirect cue.

(Refer Slide Time: 11:49)



Now, in the case of direct sighting what we are doing is when we are walking on the transect line, we are actually observing an animal. So, in the case of large sized animals, such as elephants we can directly observe the elephant, but in some cases when it is difficult to sight the animal directly we could even go for indirect cues.

(Refer Slide Time: 12:29)



Now, what do we mean by indirect cues? So, suppose you are walking on a transect line and there is a tree nearby and on this tree you saw the nest of a bird, so, this is a nest of bird. Now, if there is a nest of the bird there should be a bird who has built this nest so, because in most cases it is difficult to observe a bird directly, so, we could go for these indirect cues.

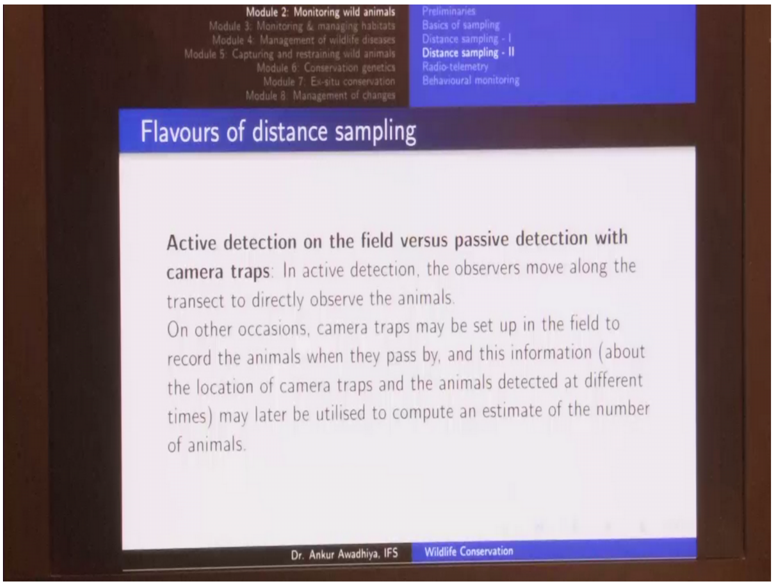
So, in the case of indirect cues you would just observe the nest we would locate it is distance and bearing from the transect line and then we would apply a correction factor. So, for instance, in the case of those birds which breed in pairs, we can just say that if we have seen a nest then there must have been a male bird and a female bird. Now, this situation is only possible when we know of the characteristics of the birds. So, for instance, in the case of weaver birds also called baya birds the male has a tendency of building a number of nests and all of those nests would be half woven. So, if you go to the forest, you would see nest hanging that is just half woven.

So, it would look like this, when a complete nest looks something like this in a cross section. So, when you see a half oven nest it means that, the male bird has built the that nest, the female bird came there and spected the nest and did not find it suitable enough. So, actually they did not pair up for that particular nest. So, any of those half nests would only mean that there is a male bird, who has built all of these nests, but we cannot say for sure that there is a pairing. But when we see a complete nest like this, then we can be completely sure that there is a male bird and a female bird that have built that nest and are residing there so, these are known as indirect cues.

Now, indirect cues could all could include the nests or in some cases we can include pellets, pellets or dung in some cases we could include even bird songs into our calculations. Now pellets are important because in the case of some animals such as the elephant; an elephant is also termed as a mega herbivore. So, it spends close to around 10 to 14 hours every day just eating. This is because, it is digestive system is not efficient enough to process the large caloric requirements that the elephant has.

So, essentially it would be eating a lot of leaves and grasses and then it would be defecating them. So, when you are moving in a transect and you see that there is a line of dung trails, then you can be sure that that it was one elephant. But then, these dung trails could be of could be fresh dung trails or these could be old drunk dunk trails; which can always be made out by looking at the texture of the dung. Also things like footprints of elephants or tracks of other animals could be used as indirect cues.

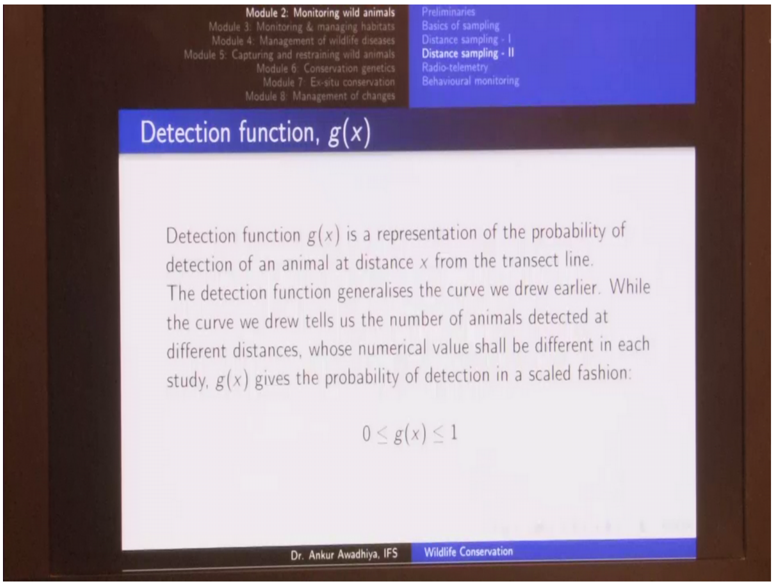
(Refer Slide Time: 06:10)



So, this is also another variant, another flavor could be active detection on the field versus passive detection with camera traps. Now in the case of active detection at the field what we are doing is that we are moving on the transect line and we are observing the animal then and there.

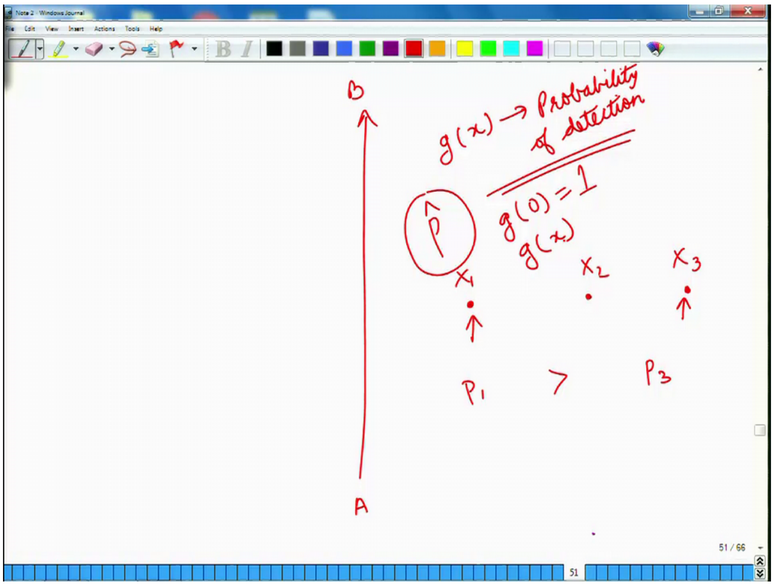
In the case of camera traps we deploy camera devices. So, such that whenever any animal moves in front of it will take a picture so, this is known as passive detection. And we can utilize those pictures also as our data points in the transects.

(Refer Slide Time: 16:45)



Now, the next concept is that offer detection function, the next concept is that of a detection function. Now detection function which is given by g of x is a representation of the probability of detection of an animal at distance x from the transect line.

(Refer Slide Time: 17:06)



So, essentially, when we are having a transect line from A to B. And suppose there are three locations of animals, so, this is X 1 X 2 and X 3. Now, the probability of detection of the animal at x 1 would be different from the probability of detection of an animal at X 3. Now, this is because, when we are having an animal that is very far away from us.

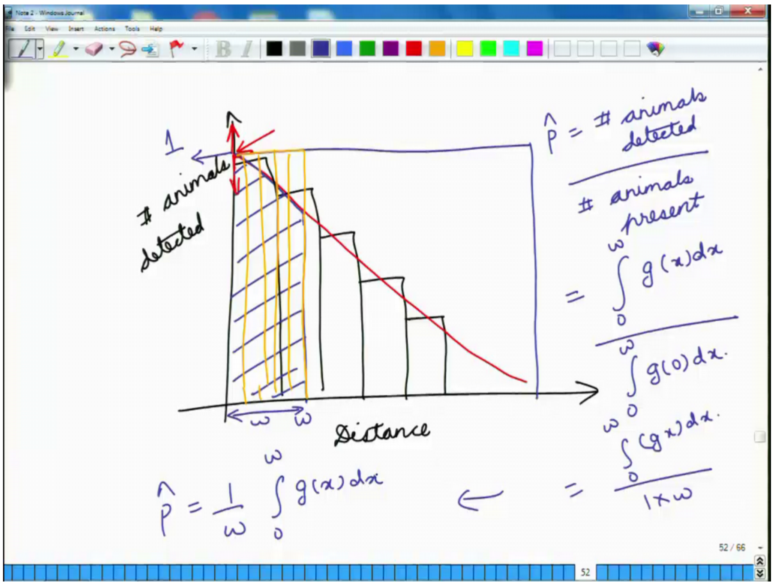
Then in those situations it is very difficult to see that animal directly and there is a very high probability that we are going to miss that animal. So, at distance X 3 will have the probability of detection say it is P 1 and here it is P 1. So, we will have we will have P 1 is greater than P 3 in most of the situations there are exceptions, but in general we can say that if the animal is close to us, we can see it more clearly, as compared to a situation in which the animal is far away from us. So, the probability of detection of an animal which is close to us, is greater than the probability of detection of an animal that is far away from us.

Now, what about the probability of a detection that is right in front of us, that is right there on the transect line. We can say that the probability of detection of this animal would be 1 or very much close to 1. Now why do we why are we be moving into this concept of g of x or the or the probability of detection. Well in the earlier class we had seen that, we had this value of p which is our detection probability and it depends on a number of variables.

So, it depends on the animal itself, it depends on the transect, it depends on where you have laid this transit with in animals, frequent this area or not. It depends on your mental state, it depends on the clothes that the observer is wearing, it depends on the food that he or she has eaten, it depends on whether those whether the observer is making any sound or not. It depends on the perfume that is being worn by the observer; it depends on whether the observer is moving singly or is moving in groups. So, there are a number of factors on which this value p depends.

Now, we wanted to make an estimate of p, now an estimate of p just because it is an estimate will depend on making some generalizations. So, one such way of deriving the value of p hat is g of x. Now with this function probability of detection, we have said that if there is an animal that is right there on our transect line, we would say that the probability of detection is 1, at distance 0 from the transect line. So, we can write it as g of 0 is 1, for any other distances we can write it as g of x.

(Refer Slide Time: 20:11)

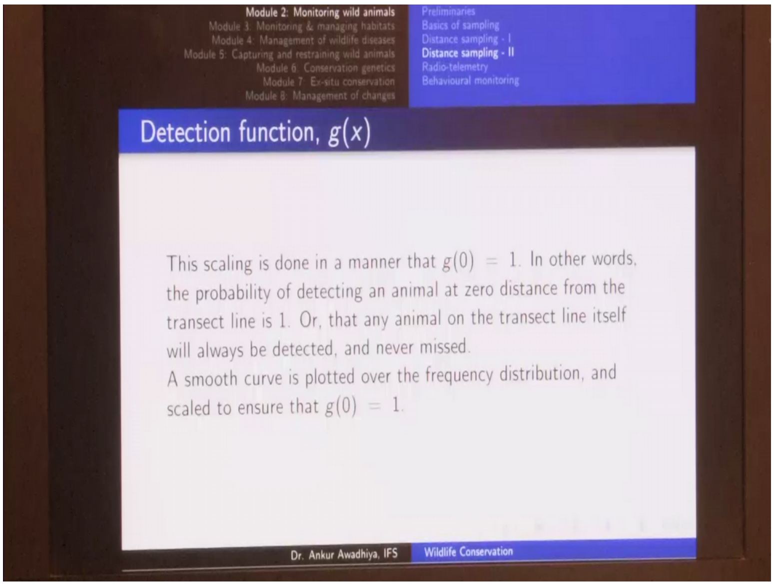


Now when we say that g of 0 is 1, what we have done here? Is that in the case of our previous curve, we had distance versus number of animals detected.

So, we had drawn all these bar graphs and then to get the value of p we had drawn curve that was moving smoothly through all of these bars. Now, in this situation, the top of the curve could be anywhere depending on situation to situation.

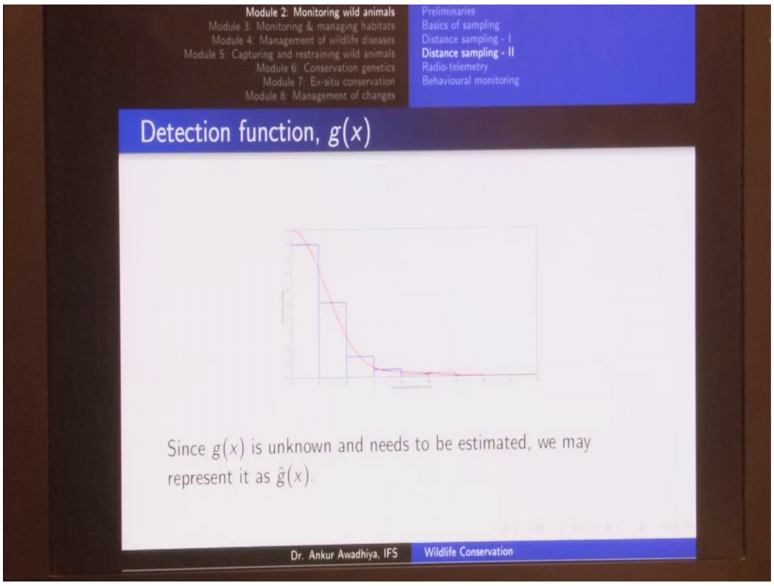
But, when we are doing, when we are writing the function as g of x we have said that g of 0 is 1. So, in that case we can scale this curve upwards or downwards so, that g of 0 is 1. So, this is one mathematical simplification that we have done.

(Refer Slide Time: 21:07)



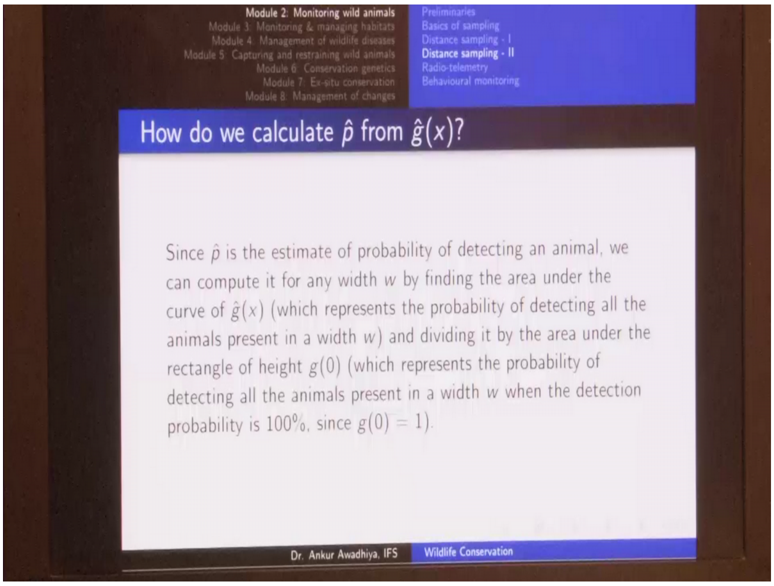
So, the scaling is done in a manner that g of 0 is 1, in other words the probability of detecting an animal at 0 distance from the transect line is 1 or that any animal on the transect line itself is always detected and is never missed. So, we draw a smooth curve over the frequency distribution and scale it to ensure that g at 0 is 1.

(Refer Slide Time: 21:26)



So, this is the curve that we have gotten. Now, once we have received this curve so the making of this curve is now simplified because, we have these bars, we made a smooth curve and then we scaled it such that the top is at one.

(Refer Slide Time: 21:48)



Now once we have this g x, how do we get the probability function over the value of p hat? Now, since p hat is the estimate of probability of detecting an animal, we can compute it for any width w by finding the area under the curve of g of x, which represents the probability of detecting all the animals present in a width of w.

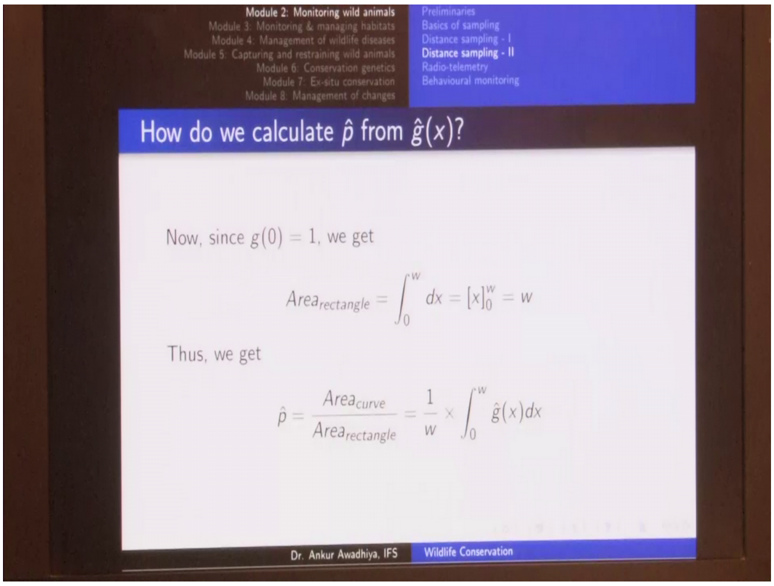
And dividing it by the area of the rectangle of height g 0, which represents the probability of detecting all the animals present in a width w. When the detection probability is 100 percent since g of 0 is 1.

So, what we are saying here Is that let us remove this. Now at any distance say w, at any distance w, the number of animals that are detected is given by the area under this curve till a width of w, so, it is given basically by this area. And the total number of animals that were there is given by the area under the rectangle of width w so, it is given by this area.

Now, since we had written that p hat is the number of animals detected divided by the number of animals present. We can write it as the number of animals detected is the area under this with these blue hashes, which is given by an integral of g x d x varying from 0 to w divided by the number of animals present. Which is given by the area under its covered by these yellow hashes, which is given by integral from 0 to w g of 0 d x.

So, essentially, if we see that the area of this rectangle is given by g of 0 which is 1 and this width is w. So, the area of the rectangle is 1 into w because, it is the length into the width and we have integral of 0 to w g x d x. So, we can also write it as p hat is given by 1 by w integral of 0 to w, g x d x.

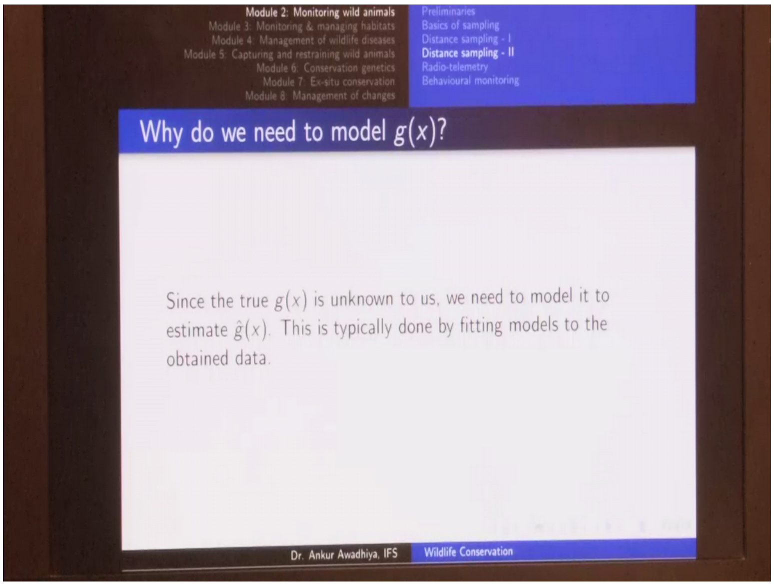
(Refer Slide Time: 24:33)



So, which is this equation, p hat is the area of the curve divided by the area of the rectangle is 1 by w integral from 0 to w g x d x. Now here again we have a hat on top of g because, this again is an estimate, we have not yet figured it out completely.

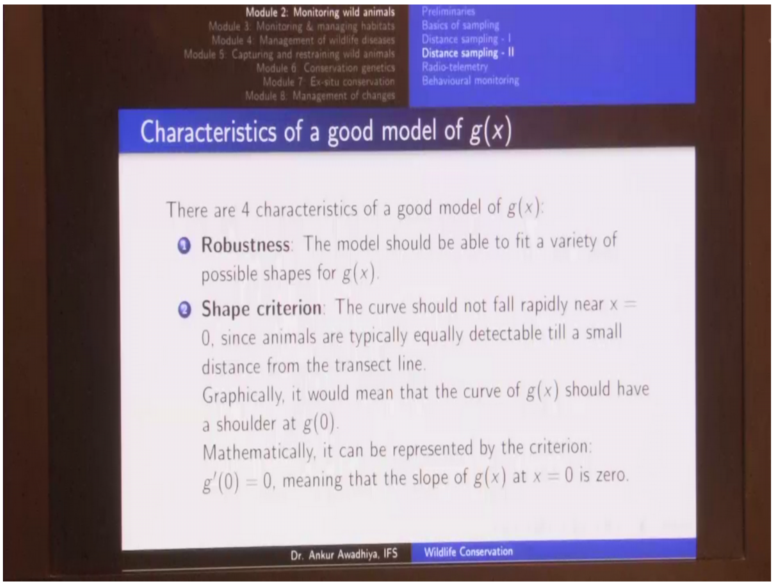
But, from here we know that once we have figured out an equation for g of x we can compute p hat. So, even though p hat depends on a number of variables, now we are coming close to a way of mathematically getting to the value.

(Refer Slide Time: 25:07)



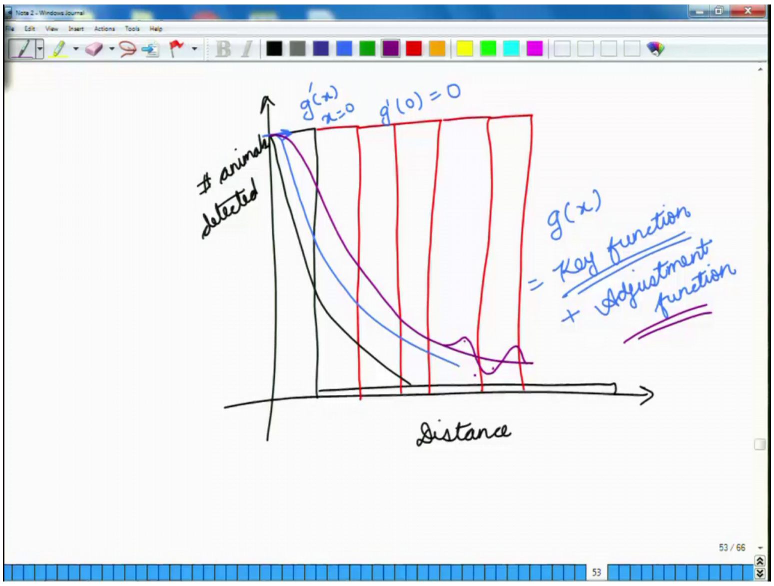
So, when we perform this computation on a computer, we need to it is, we need to model it in some way.

(Refer Slide Time: 25:23)



So, when we use a model of g x there are four criteria that we use for a good model of g x. The first criteria is that of robustness, so, robustness means that, the model should be able to fit a variety of possible shapes of g x.

(Refer Slide Time: 25:36)



Which means that in certain situations we can have that we have a good detection at so, here we have distance versus the number of animals detected. Now in certain situations we will have a situation in which we have a very good detection very close to the transect line and then practically 0 detection everywhere else.

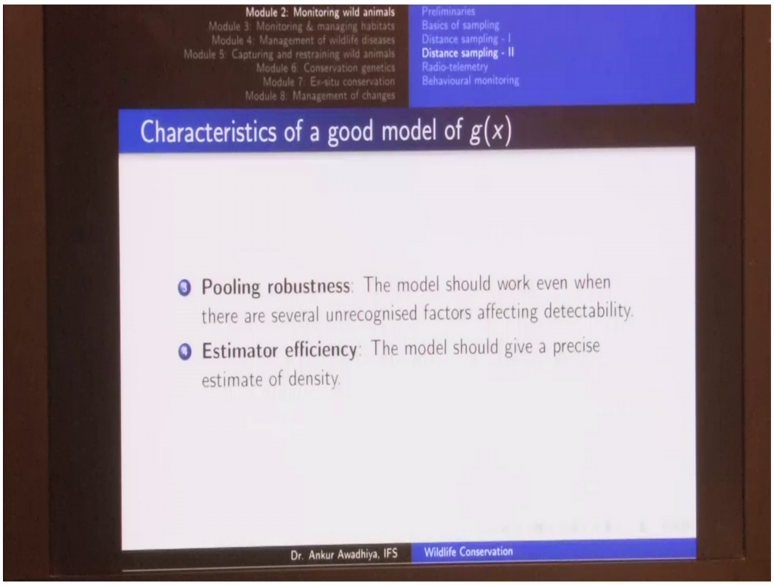
So, a situation would mean a something like when we are walking on a patch of grassland and there is a very small species. So, if it comes on right on the transect line we will be able to see it giving a very good detection. And as soon as it moves into the grasses we will miss it giving a detection of nearly 0. Now another situation could be where we have a detection which is roughly the same everywhere.

Now, such a detection is possible say when we are looking at elephants in a grassland. So, the elephant being a large species we can see it even at a very great distance so, it is it is detection would be nearly equal everywhere. Now, when we are choosing a detection function; it should be such that it explains both of these extremes and anything else that also can come in between so, that is the first criterion.

Now, the second criterion is that of a shape criteria, now a shape criterion means that, when we are saying that when an animal is right there on the transit line. And we are able to see this, if the animal moves a bit from this transect line, do we see it or not? Now in both situations if you are able to see it right there on the transect line, even if it moves a bit you will be able to see it.

Which means, that we are when we are selecting any detection function g of x, then it should have a slight shoulder right near the transect line. So, a shoulder right near the transect line, would mean that we will not be having a detection function that just moves like this, but, we will prefer a detection function that has a slight shoulder here and then moves down. Now, this criterion if we put it mathematically, we will see that the slope at 0 that is g prime x at x equal to 0 or also g prime of 0. So, it should be as close to 0 as possible, but it will be a finite value so, that we have a small shoulder at that point.

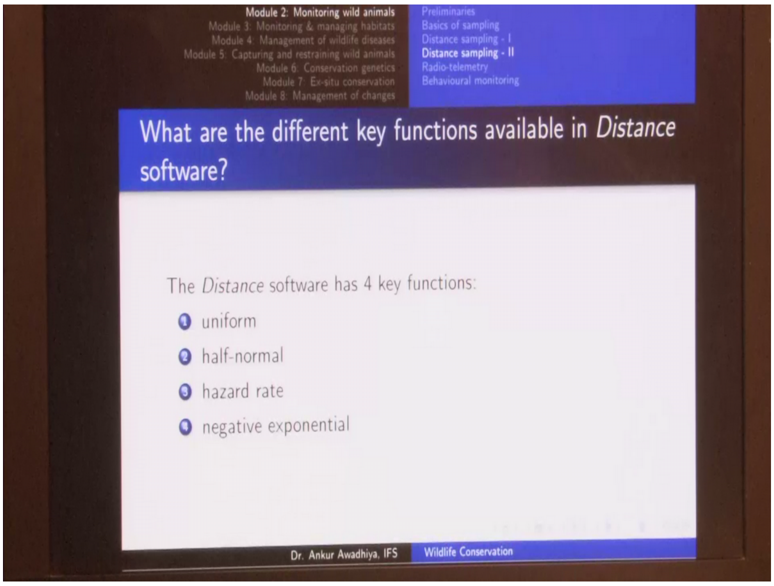
(Refer Slide Time: 28:15)



The third criterion is that of pooling robustness, now pooling robustness means that the model should work, even when there are several unrecognized factors that are affecting detectability. So, as we had seen in the last lecture there are a number of unrecognized factors and our model should be able to incorporate most of them. The forces estimated efficiency, which means that the results that we are getting should be precise, as precise as possible.

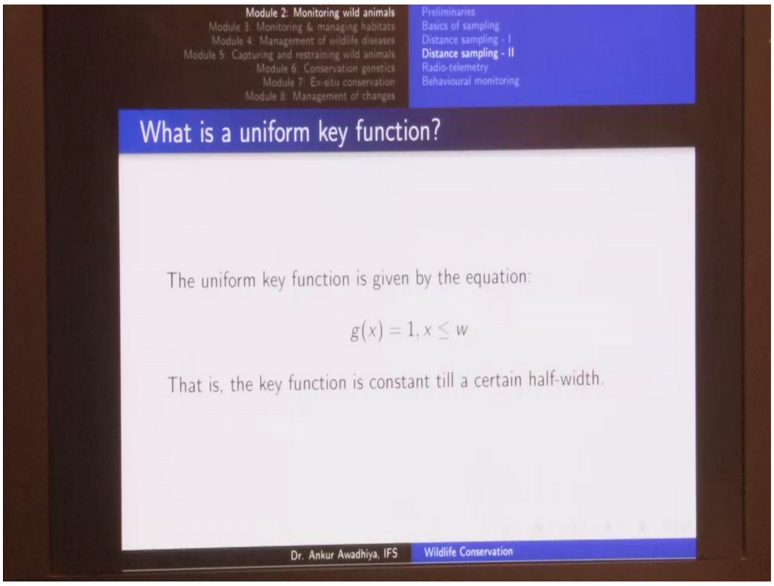
So, when we write g of x, in a general form we can write it as g of x is given by a key function plus an adjustment function. What this means is? That the key function refers to a shape that we have chosen for our g of x so, if this is our shape for the g of x. And if there are some points above and below so, the key function would give us the general shape of the g of x. And adjustment functions would be used to modulate our curve so, that all these points that are coming out of the curve are also taken care of.

(Refer Slide Time: 29:37)



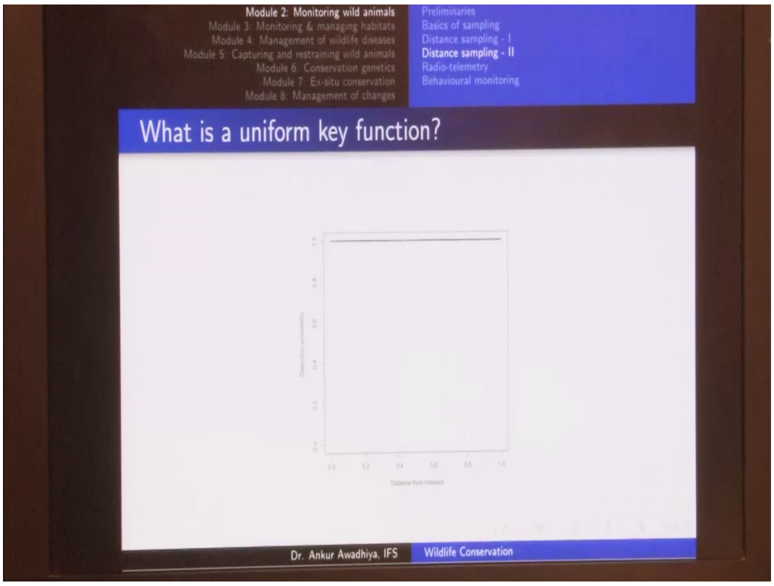
Now, there are four key functions that we normally use, so, key function is telling us the general shape of g of x. So, the four functions that are commonly used are uniform function, half normal, hazard rate and negative exponential.

(Refer Slide Time: 29:53)



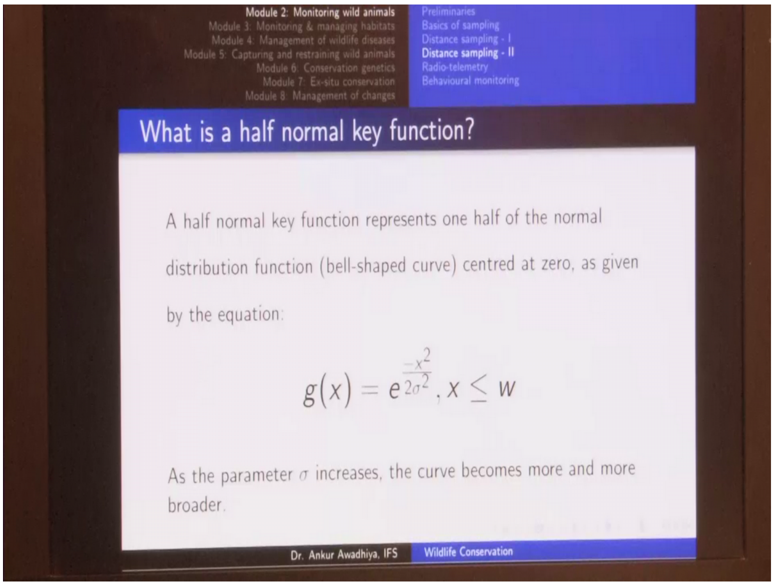
Now, what do we mean by a uniform key function? It is very easy to remember a uniform key function by taking the example of elephants in a grassland. So, at very large distances from us we will be able to see the elephants. So, our detection probability is going to be the same everywhere.

(Refer Slide Time: 30:10)

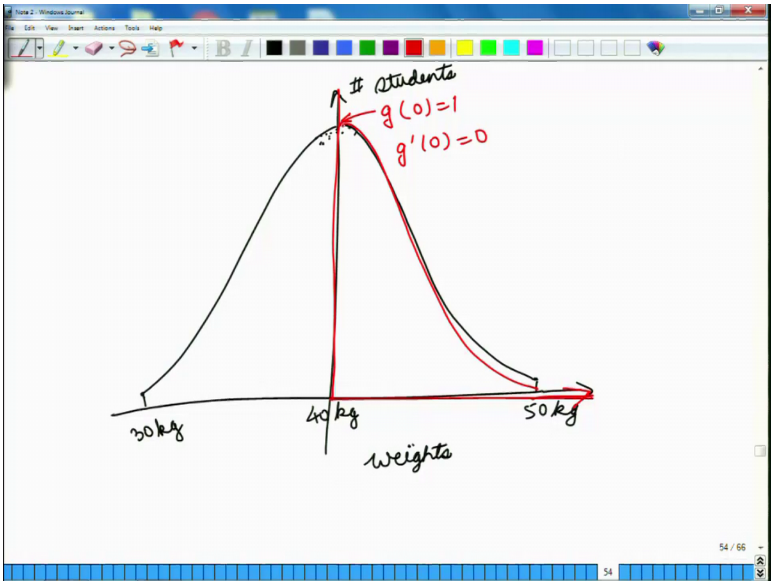


So, this is how uniform key function looks like so, here we have that g at 0 is 1 and even at large distances g of x is 1 so, this is a uniform key function.

(Refer Slide Time: 30:25)



(Refer Slide Time: 30:35)



The second key function is given by the half normal key function, now what is a half normal key function? Now, a normal curve is what we observe in most situations as a bell shaped curve so this is a bell shaped curve.

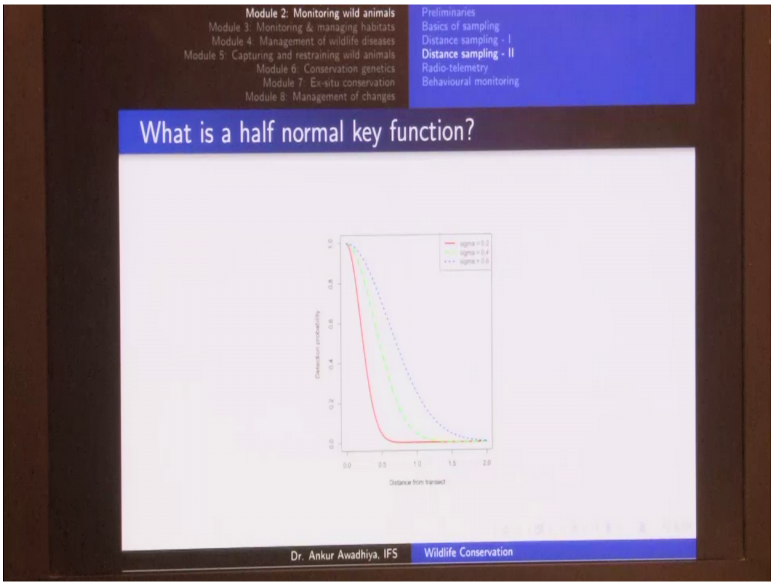
Now a bell shaped curve is the most common curve that is seen in biological samples. So, for instance, consider the heights of students in a class or say the weights of different students in a class. So, there would be a midpoint at which most of the weights hover. And then, if we go towards the extremes then there would be less and less number of people that are having those weights.

So, for instance in a class where say the midpoint is 40 kgs and the extreme number of people that have the weights say 50 kg and 30 kg. So, we would find that most of the people have weights that are very close to 40 kg so, which is and so, this is the number of students and on the x axis we have the weights.

So, we will find that most of the students hover like this in this region so they have weights somewhere say 40 kgs, 40.2 kgs, 39.5 kgs and so on. And then, if we look at the extreme levels so, there could be a student that has a weight of 50 kg, there could be say a few students that have the weight of 30 kgs but, in general we will see a bell shaped distribution. Now a half normal takes one half of this part so, this is a half normal this shape. So, if we take such a shape for our detection function, we have that at the top we have g at 0 is 1. We also have that g prime of 0 is 0, which means that we have a small shoulder here and this is another curve that we can very easily use.

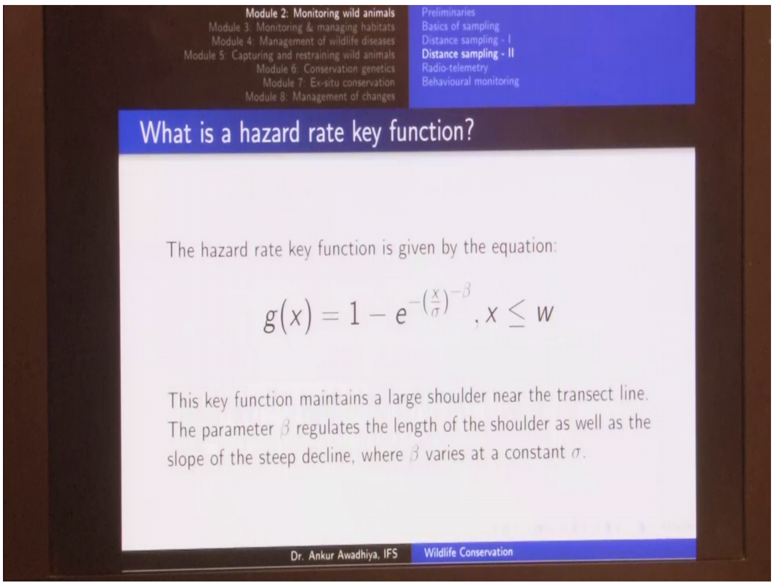
Now, the equation of the curve is given as g x is e to the power minus x square by 2 sigma square, when x is less than equal to w. Now, for this particular course it is not essential to remember all these formulae, but just a general shape would suffice.

(Refer Slide Time: 32:50)



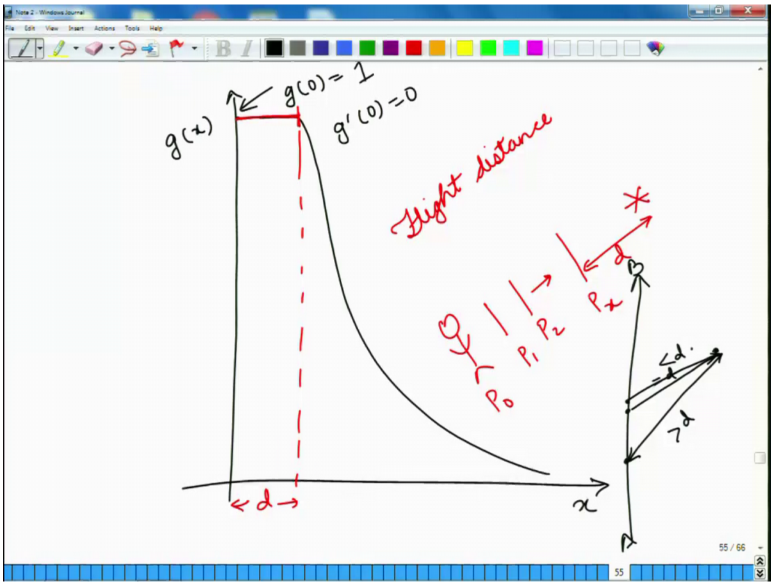
So, basically this is our half normal so, you can have different kinds of half bell shaped curves depending on the values of.

(Refer Slide Time: 33:00)



Now, the next function for g of x the next key function is called the hazard rate function.

(Refer Slide Time: 33:06)



Now, we can understand a hazard rate function by looking at its shape. So, the hazard rate function has a broader shoulder as compared to half normal and then it goes down with distances x. What do we mean by this shape? It means that at g of 0 we have it as 1 and g prime of 0 is 0.

So, we are having that at the top right on the transect line we have a detection probability that is 1, and also we have a shoulder here, but how do we understand this curve intuitively. Let us consider an animal that shows a phenomenon called flight distance. Now, what do we mean by flight distance? So, suppose you have an animal here and suppose this is your position, now if the animal has observed you and it freezes. So, when we say freezing, it has observed you and just to avoid detection it just stands still there so, it will not move.

So, now as you go on approaching this animal there would be so, you came to say this point. So, this is your point p 0, then you reached a point called p 1. Even at p 1 this animal is just standing there it is it has just frozen at its place. So, you go a bit closer at distance p 2 at point p 2.

So, even then it is freezing, but then as you go on increasing your nearness to the animal, there would be a certain point say p x at which this animal now thinks that you have observed this animal and you are coming to grab this animal. So, what it will do is it will run at that point.

So, this distance of d till the point where the animal is able to tolerate you is known as the flight distance. So, if you are anywhere greater than the flight distance, the animal is going to freeze, if you are anywhere within this flight distance, the animal is going to run away.

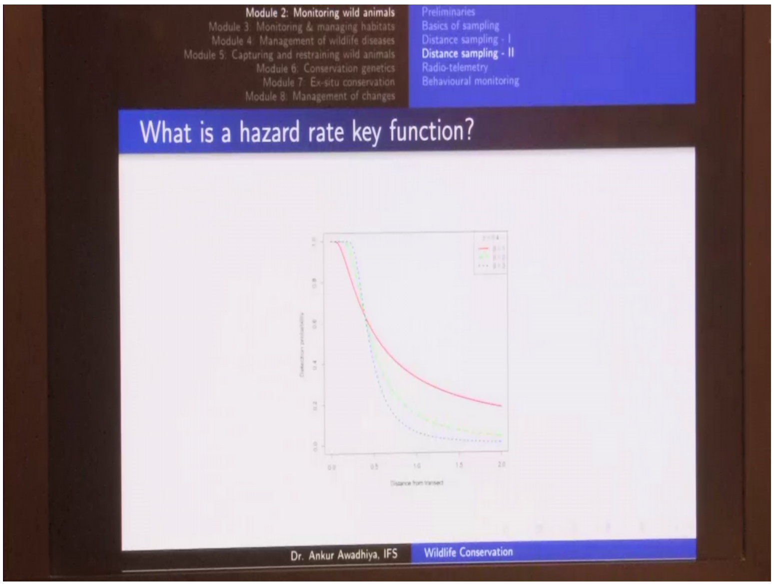
Now, how it affects our detection of the animal is that when the animal is standing frozen, right there on the ground, when it is not moving and it is very well camouflaged. So, we will not be able to detect that animation, but when this when we have approached this animal. And when we have crossed it is threshold of the flight distance this animal will run away. Now, when an animal moves it is very easy to detect that animal. Now coming back to our curve, it shows that when we are at a distance of d from the animal, this animal runs away which facilitates it is detection so we have this broad shoulder on the top.

So, till our distance d we have a very good detection and nearly complete detection. So, you we have g of x is equal to 1 till the tailored distance from the animal is less than or equal to d, but right after that, as soon as we have crossed this distance. So, now, the animal is comfortable, now this animal is not going to run away. So, we will not be able to detect this animal. So, when you are walking on a transect line, if the animal is at a certain distance and if this distance is greater than the flight distance, then the animal will just stay frozen there and it will not move.

So, we will not be able to detect this animal. But as soon as we have a situation when this animal is closer to the to the transect line, then the flight distance. So, basically what we are saying here is that when we are moving on our transect line A to B here we have this animal and we are at this position.

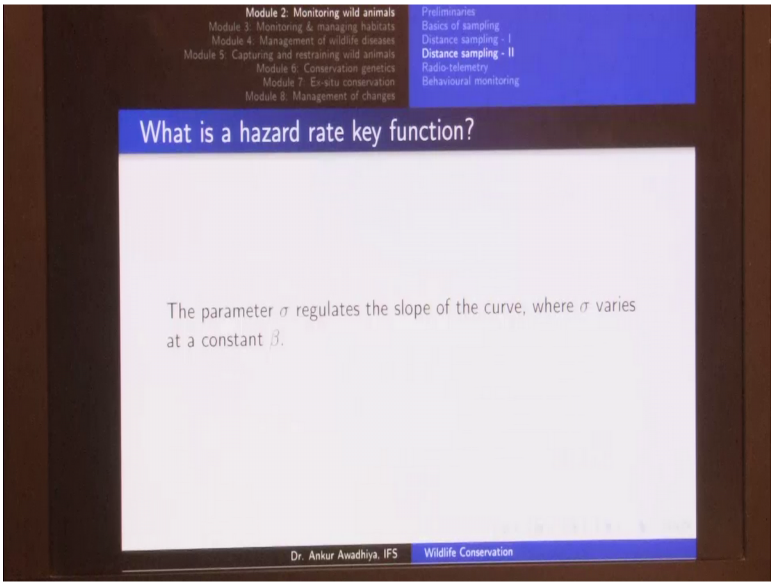
So, at this position, this distance is greater than d so, this animal is just standing there. But then when we came to this point now our distance from the animal is equal to d. So, even here the animal is right on the threshold, but we take one more step on our transect and our distance becomes less than d and this animal feels away. So, essentially, we will be able to detect that animal till a distance of d which is shown by our curve.

(Refer Slide Time: 37:37)



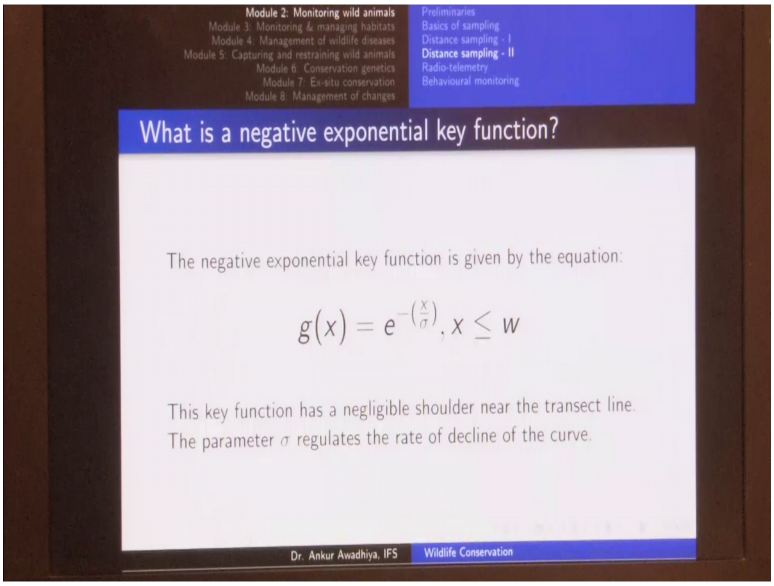
So, this is our general shape of the hazard rate key function, so, we have a shoulder on the top and then this curve moves down.

(Refer Slide Time: 37:49)



Now, it depends on two parameters so, the parameter sigma regulates the slope of the curve, when sigma varies at a constant of beta. So, this is how it will look so we have kept beta constant at 1 and we are varying sigma so, this is how the shapes are varying.

(Refer Slide Time: 38:03)



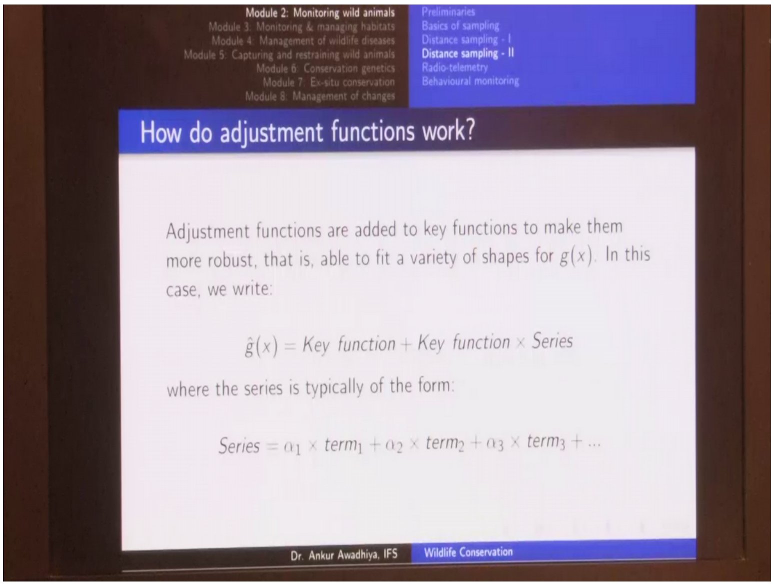
And similarly so, again for this course it is not essential to remember the formulae, but just a general shape of the curve in your minds will suffice. Now, the next function the next g x is given by the negative exponential key function. Which has this formula g of x is exponential of minus of x by sigma.

Again looking at a common shape, we have that at this point we are able to detect the animal, but right after it the probability goes very close to 0. Now, it is important to note here, that in the case of a negative exponential key function it does not have a shoulder on the top. So, essentially our criteria of g prime of x at x is equal to 0 should be 0 is not fulfilled by this negative exponential key function, but it also becomes essential for those situations in which you are able to detect your animal only when it is right there on the transect line.

So, this happens in situations when we are considering say reptiles so, if there is a snake and snakes are generally highly camouflaged animals because, they are predators. So, if there is a snake that is right there on you transect line you will be able to see it, but as soon as it goes into the grasses there is very little chance that you will be able to see it.

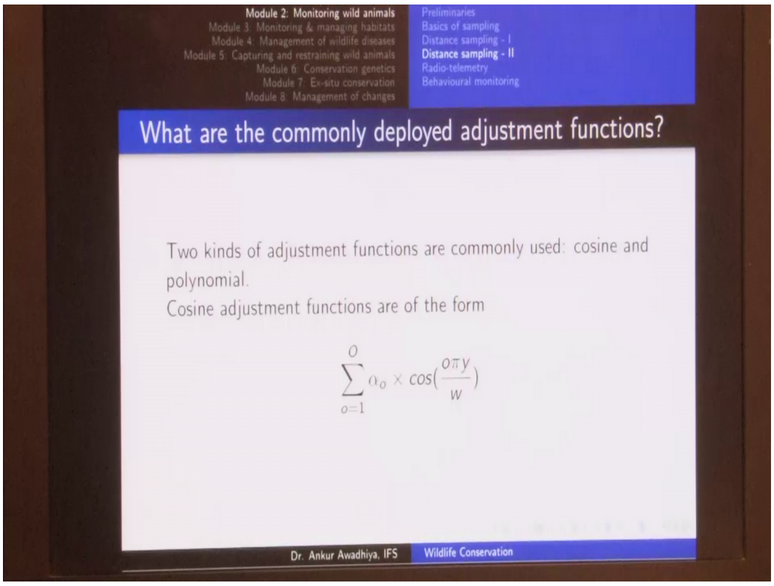
So, this curve essentially shows that phenomenon so, right there on the transect line you have a good detection probability, but then it drops down very fast. So, these were the key functions.

(Refer Slide Time: 39:40)



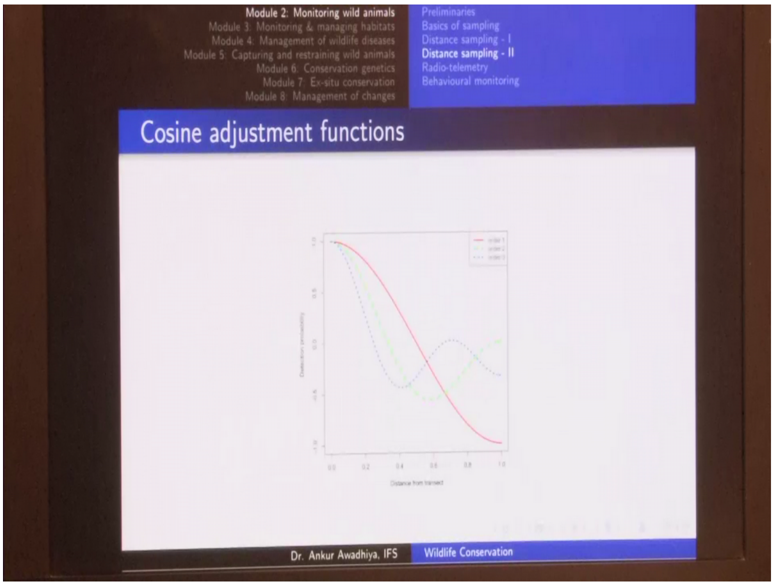
Next we have the adjustment functions so; we have the g of x is given by a key function, plus an adjustment function. So, your adjustment function can also be written as a key function multiplied by something.

(Refer Slide Time: 40:01)

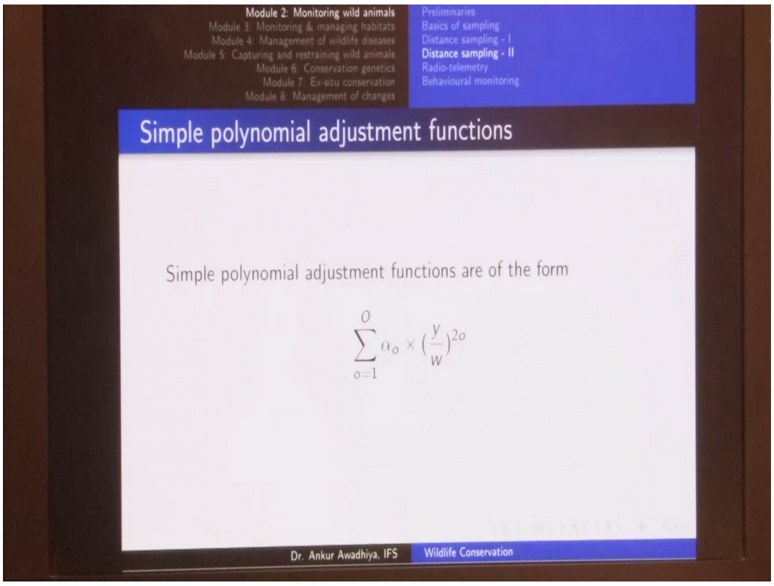


So, the adjustment functions are generally used as in two ways so, two commonly deployed adjustment functions are the cosine function again the formula is not important, but the shape is important.

(Refer Slide Time: 40:04)

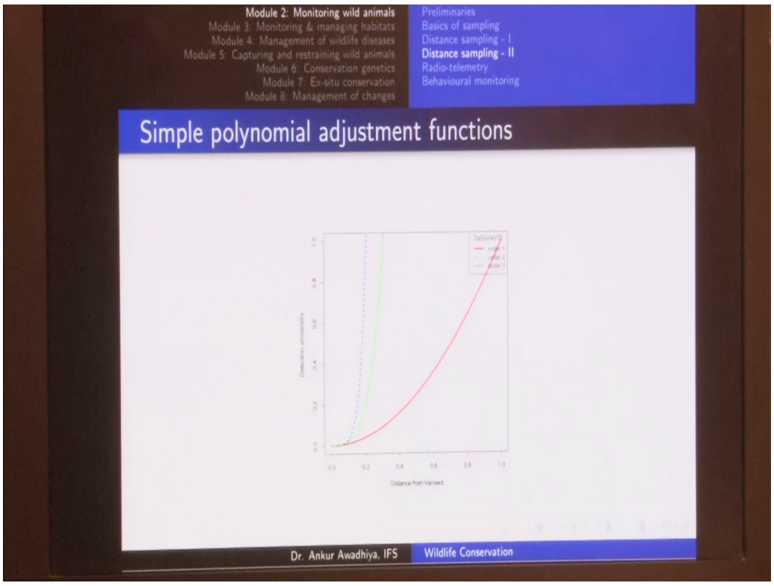


(Refer Slide Time: 40:16)



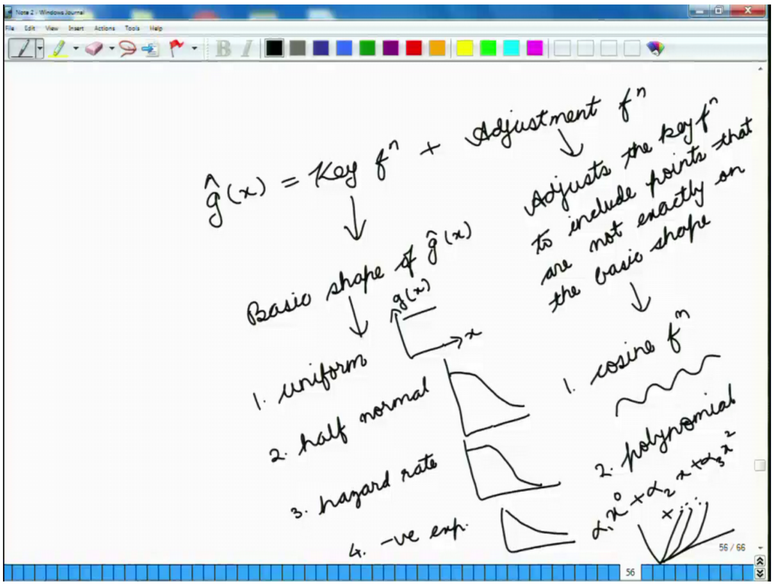
So, in the case of a cosine adjustment function, we will have these wavy lines which represent our cos theta values. And the other one is a simple polynomial adjustment function.

(Refer Slide Time: 40:19)



So, this is our simple polynomial adjustment function so, basically it could be something like y is equal to x y is equal to x square y is equal to x cube or any combination of these.

(Refer Slide Time: 40:34)

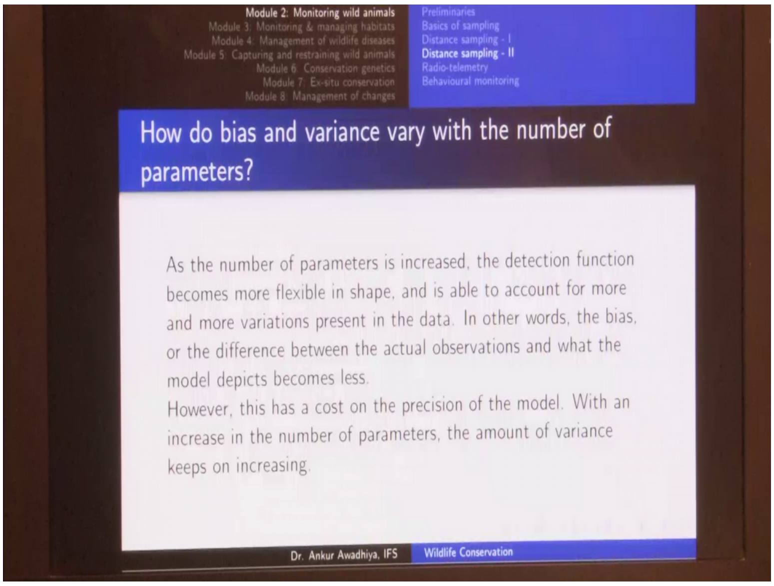


So, essentially what we have said is that the g of x is a key function plus n adjustment function. Now, key function is the basic shape of g of x. And there are 4 basic shapes that we have seen, one is uniform; which goes like a straight line, the second is a half normal, which is half of our bell shaped curve. The third is a hazard rate, which essentially is a shoulder followed by going back to 0.

And the fourth is a negative exponential, which is just going to 0 right after your the value of 0. So, here we have in all these curves we have the g of x versus the x. So, key function is explaining the basic shape of the curve. Adjustment function, adjusts the key function to include points that are not exactly on the basic shape.

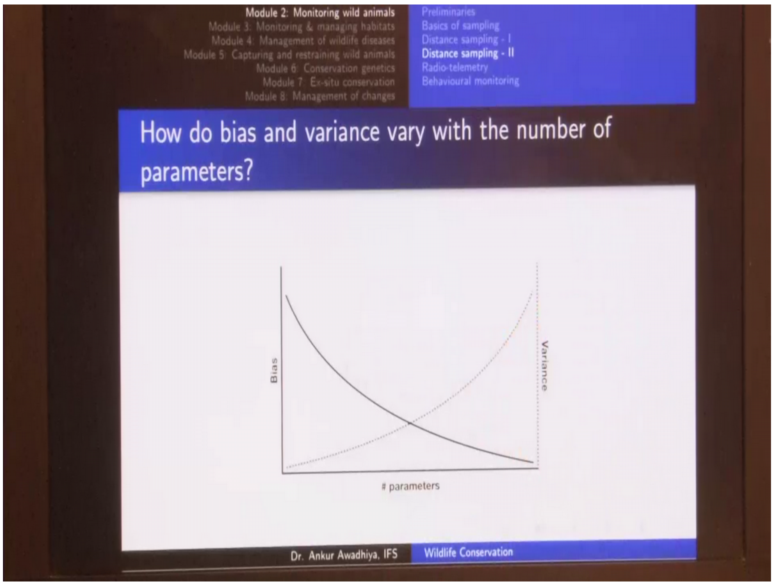
And here you saw two adjustment functions, one is the cosine function which moves like this and the second is a polynomial, which could be x to the power for 1 x to the power 0 plus alpha 2 x plus alpha 3 x square plus so on. So, essentially, we are saying that our function here is either y is equal to x or x square or x cube and so on. So, basically these are the two key things to remember, key function and adjustment function.

(Refer Slide Time: 42:26)



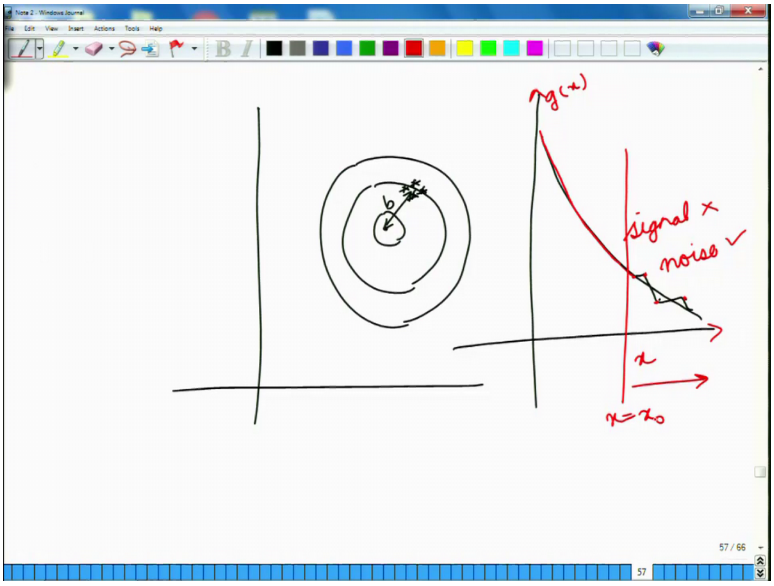
Now, how many adjusting functions do we use? So, here we have this concept of bias and variance.

(Refer Slide Time: 43:03)



So, essentially, if we consider the number of parameters so, if there are more number of parameters, what happens? Now, bias is the difference between the readings and the actual reading so, if we remember our case of the bull's eye shooter.

(Refer Slide Time: 43:17)

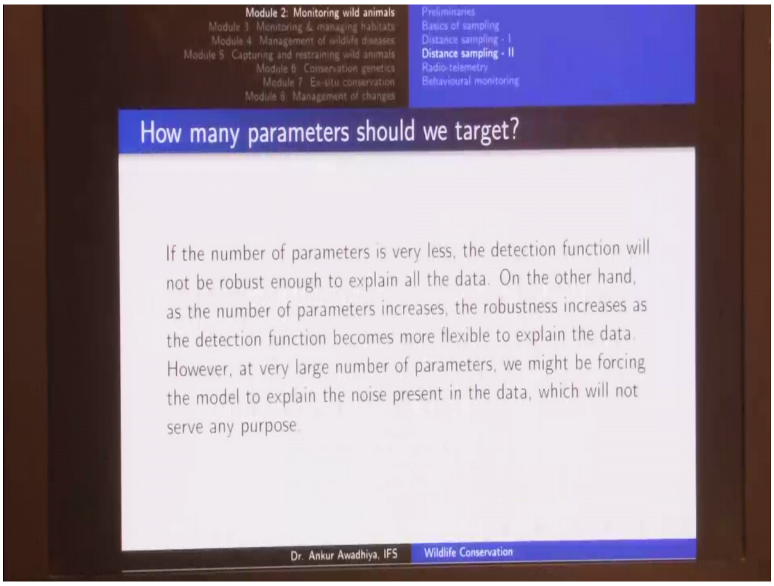


So, this is our bulls eye and the shooter had all the points here so, this distance is our bias b. Now if we are including more and more terms or more and more parameters into our equations, the bias reduces, why does the bias reduce? Because in place of having our so, this was our model of g of x, but actually our points were lying somewhere here.

So, we included some adjustment functions so, that all of these points are also covered. So, in that case the level of bias reduces, but at the same time, if you look at variance the variance increases because what we are doing in this case, is that all these points that we were having. So, these points that we were considering, these points might not be representing the actual signal, but might actually be representing the noise values.

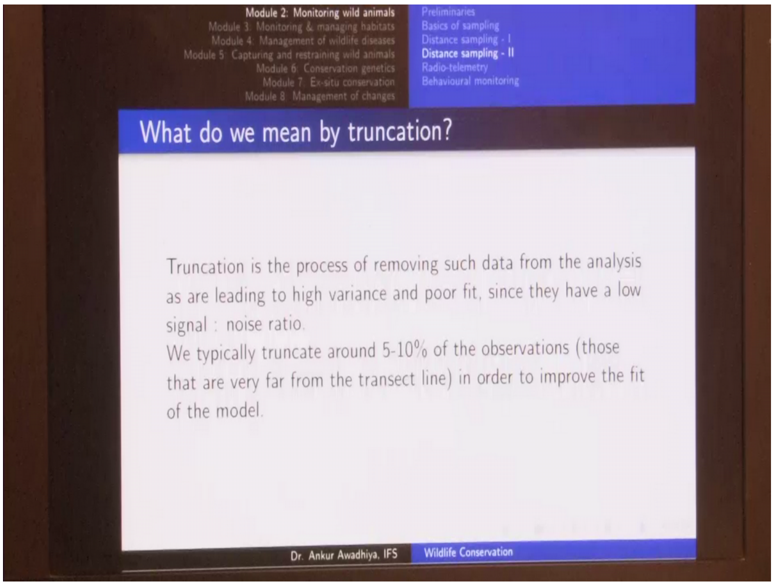
So, noise values means, that that these are those data points that we are not very sure of. So, in that case as we increase the number of parameters, the amount of variance goes on increasing. So, the next question is how many parameters should we take? It is always best to take those number of parameters for which your bias is less, but at the same time your variance is also less. So, essentially a mid number of parameters so, do not over adjust your g of x, do not under adjust just your g of x take a middle point.

(Refer Slide Time: 45:07)



So, we should target middle number of parameters.

(Refer Slide Time: 45:13)



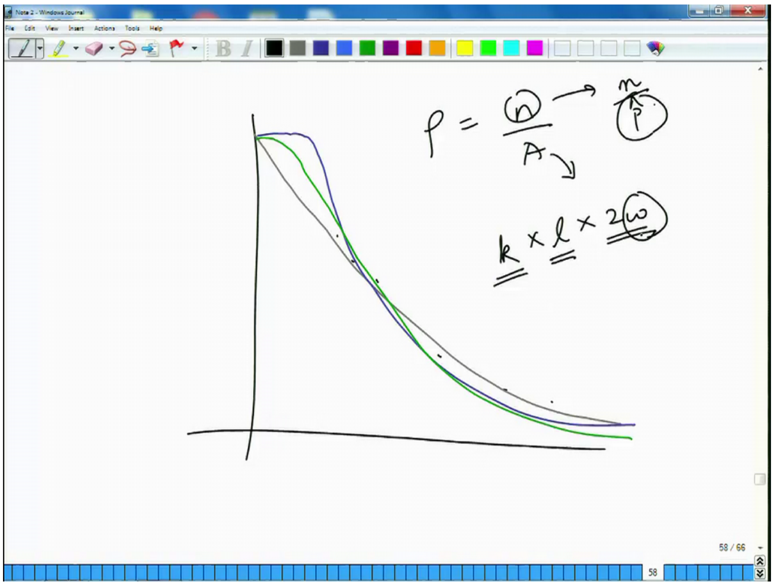
Now, another thing is truncation, now in the case of truncation what we have is that when we consider the value of g of x versus x, we can also say that for these values, we are very much sure that of having an actual observation. So, essentially, when you are moving on a transect line and there are animals that were very close to your transect line.

So, you are very sure that you have seen these animal, but then consider an animal that was very far away. So, now, you are not sure whether what you saw was actually an animal or whether it was say a rock of the same color. So, essentially at very large distances was it an elephant or was it a black colored rock you are not very sure.

So, in those cases to improve our accuracy we can just say that we are going to disregard all the values, that we got after a value of x is equal to x 0. So, this value of x 0 can be chosen by us and we can say that any values that were greater than then this distance we are going to truncate it out. So, truncate it out is it means, that we are going to leave it out in our calculations.

So, we typically truncate around 5-10 percent of the observations. So, those observations that are very far away from the transect line, in order to improve the fit of the model without resorting to a very large number of parameters. Now once we have selected a model and once we have tried to fit our model.

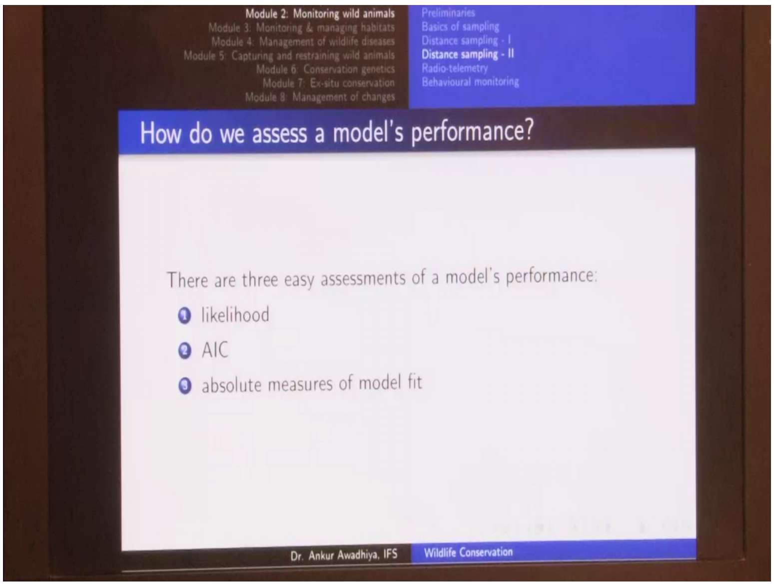
(Refer Slide Time: 46:49)



So, essentially when we have some points, we could fit a model say like this. A negative exponential model or we could say that we are going to fit a hazard rate model something like this or we could say that we are going to fit a half normal model something like this.

Now, you can see that none of these models is actually catering to all the points that are there on the our data. So, how do we assess which model to take?

(Refer Slide Time: 47:21)

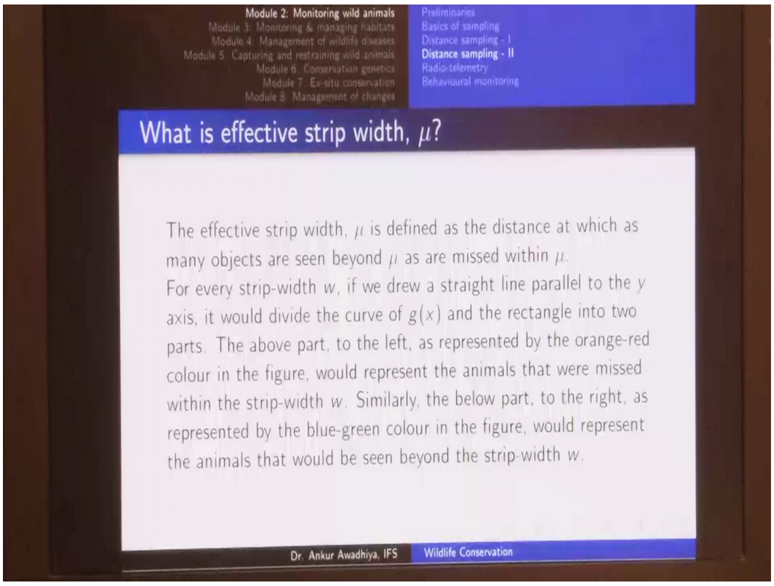


So, essentially, there are three easy ways of assessing a models performance, which go by the name of likelihood AIC which is a Akaike information criterion and absolute measures of model fit.

So, for the purpose of this course it is not essential to get into the details, but basically, we should only understand that it is mathematically possible to understand. Which of these different key functions and different adjustment functions is mathematically the function which is using the least number of parameters. And is able to explain our data points to the highest degree, which is actually what we want. What we are looking after is just a value of g of x from which we can get a value of p hat and we can have different g of xs as we can see on the screen.

So, we can have different g xs so, we can have a negative exponential, we can have a half normal, we can have a hazard rate and so on. But two of these should be taken we can compute it mathematically so; our computer can give us the results.

(Refer Slide Time: 48:24)

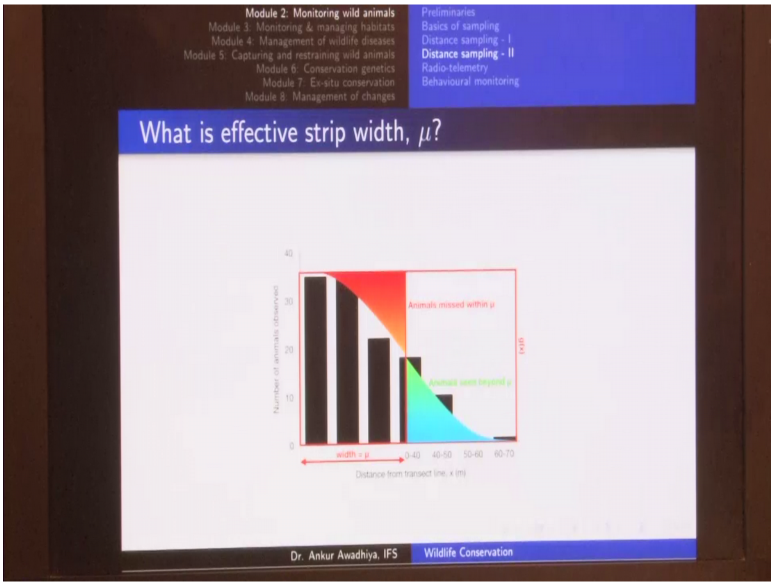


Another concept is that of the effective step width, so, once we have decided our g of x. The next thing that we want to compute the density, is the density is given by the number of animals, that were there divided by the area. And the area is given by the length of the transect multiplied by price of the half width.

So, we can also multiply it by k which is the number of transects that we took and the n can also be modified as n by p hat. So, we can compute this p hat with our g of x by 1 by w integral of the g x, but and we know the number of transects that we have walked, we know the lengths that we have walked.

Next thing to figure out is w so, can we have a mathematical function through which we can find out the w. So, here comes the concept of the effective strip width. This the effective strip width mu is defined as the distance, at which as many objects seen beyond mu are missed within mu.

(Refer Slide Time: 49:35)



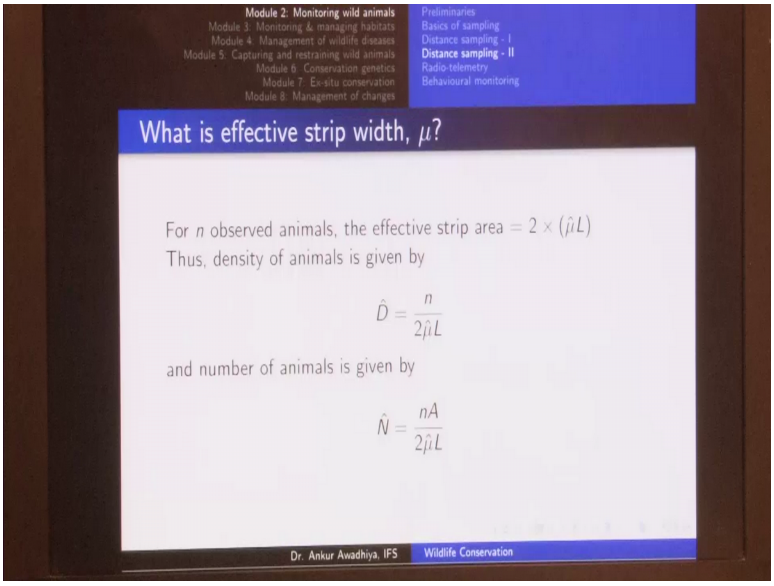
So, what we mean here, is that in the case of our curve of g of x we can always draw a line. So, at this distance w so, anything that is that is below our curve of g of x is what we have detected. So, this portion which lies within the value of g 0 which is 1 and g of x is the number of animals that we have missed out at distance w.

On the other hand, the number of animals that are seen beyond this width of w is given by the area of the curve below g of x so, which is given by the blue shade here. So, at every distance w there will be some animals that have been missed out within this distance and there will be some animals that were seen outside of this distance.

Now, as we go on increasing this w will reach a point at which the number of animals that were missed within this distance given by the red color, is the same as the number of animals that were seen beyond this distance.

So, both the curves have an equal area. So, at this distance, we say that the width is given by mu so, mu is the effective strip width. So, why do we need mu here? Because coming back to the screen, we require a value of w and when we use this value of mu as w, we are we will get our results.

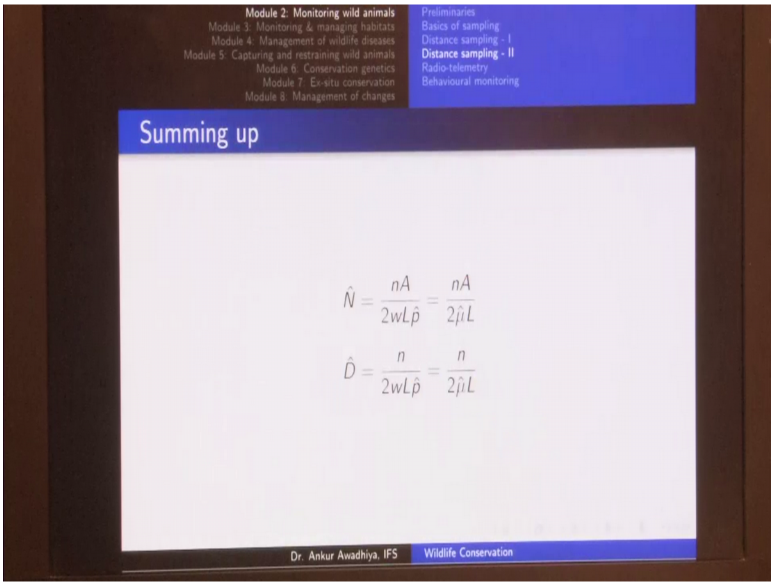
(Refer Slide Time: 51:07)



So, essentially, for n observed animals the effective strip area is given by twice of mu. So, which is twice of the half width multiplied by the length, now here this capital L because, it is small l into k. So, this is the effort or the total distance that we have moved.

So, the density of animals will be given by D hat is equal to n divided by this effective strip width to mu of L and the number of animals will be given by this D hat into A. So, it is n into capital A divided by twice of mu hat into L. Now all these formulae are not important for this course, but what we need to remember here, is that it is possible to find out a value of g of x from which we can get a value of p hat. And it is possible to compute your mu hat or the effective strip width also from the g hat from the g x curve. So, once we have done that, we can use this value of mu into our equation to get the density figures and the estimates of the animals within our forest.

(Refer Slide Time: 52:15)



So, in these two lectures on distance sampling, we saw how we modulate our equations for the density. So, density essentially means, number of animals per unit area. Now, we saw that we are missing out some animals so; we included the concept of p hat. So, we modulated our value of n by n by p hat to incorporate those animals that we missed out.

Now, we miss out the animals, because of a number of factors. So, it could be anything and everything that can get that can influence our detection of the animals. Now to compute a value of p hat, we went through this exercise of finding out a curve g of x which had some criteria that g at 0 is 1. And there is a small shoulder at distance x is equal to 0 and from that we can compute our value of p hat or similarly, we can compute a value of mu hat.

So, once we have done that, we get our values of N and D, which is the estimate of the number of animals in the forest or an estimate of the density of animals in the forest. So, distant sampling is one very important technique, that we are using these days, to estimate the number of animals by incorporating all the factors or most of the factors that can influence our detection of the animals. So, that is all for today.

Thank you for your attention jai hind.