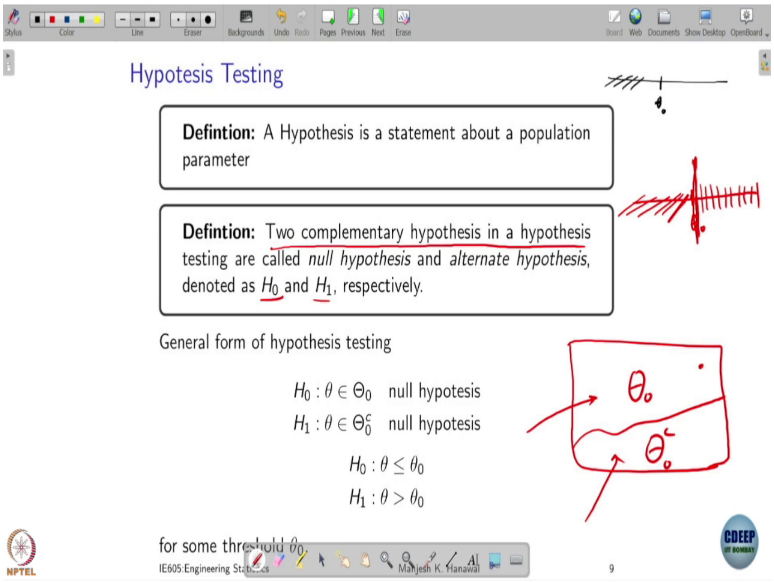
Engineering Statistics
Professor Manjesh Hanawal
Department of Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay
Lecture 42
Hypothesis Testing, Bayes Test

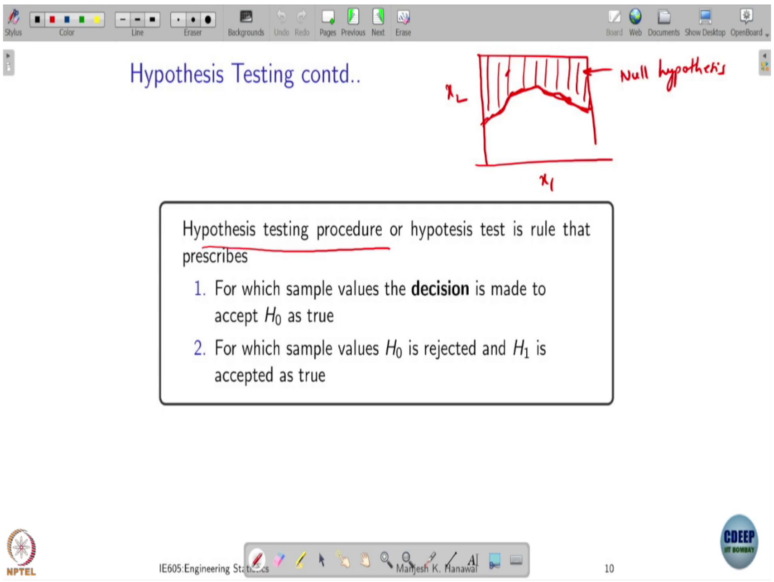
So, we started talking about hypothesis testing. And we said that hypothesis is a statement about a population parameter.

(Refer Slide Time: 00:29)



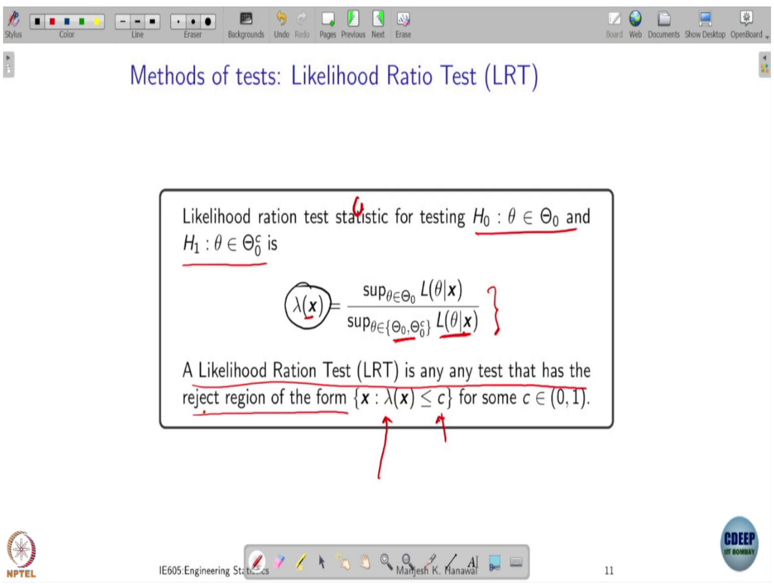
And we said in hypothesis there are actually two hypothesis, H0 and H1, which are complementary. And the H0 hypothesis is basically a statement like my parameter theta belongs to some set big theta 0. And that we are calling it as null hypothesis. And alternative hypothesis is my theta belongs to the complement class, which I am calling it as null hypothesis. And if my theta's are like real line, I may put a partition somewhere. Like if my theta space is like a real line, I may say okay, I am put this as my theta 0 and ask whether I this is below this or above this.

(Refer Slide Time: 01:24)



So, in hypothesis testing, we are going to come up with a rule that tells whether my hypothesis need to be accepted or rejected. So, the rule is going to give me tell me how to make decisions.

(Refer Slide Time: 01:46)

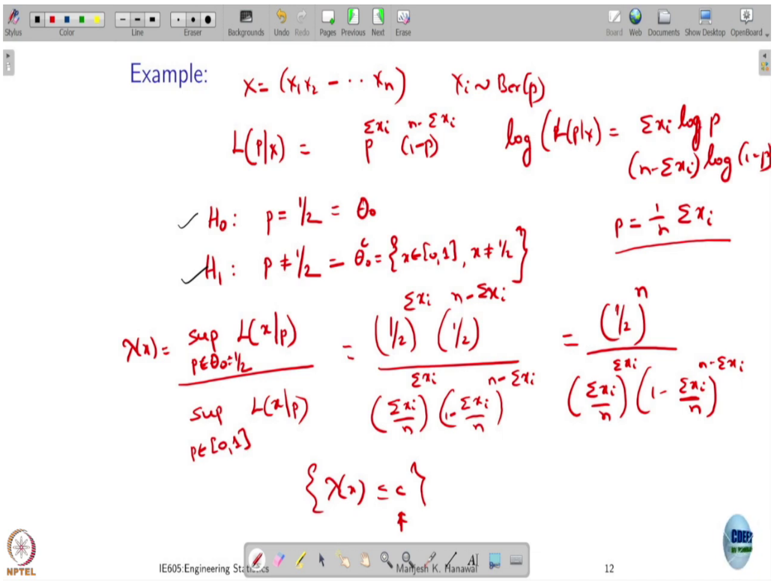


And one of the possible decisions criteria we came up, is based on likelihood ratio test, which we call LRT. So, LRT is defined as a ratio of 2 quantities, which are defined in terms of your likelihood function. The numerator is optimizing your likelihood function over theta naught, which is basically the parameters belonging to your null hypothesis. And that denominator is optimizing over the entire parameter space, which is theta 0 as well as theta 0 complement.

So, this ratio whatever we are going to get for a given samples x be denoted as L of x sorry lambda of x.

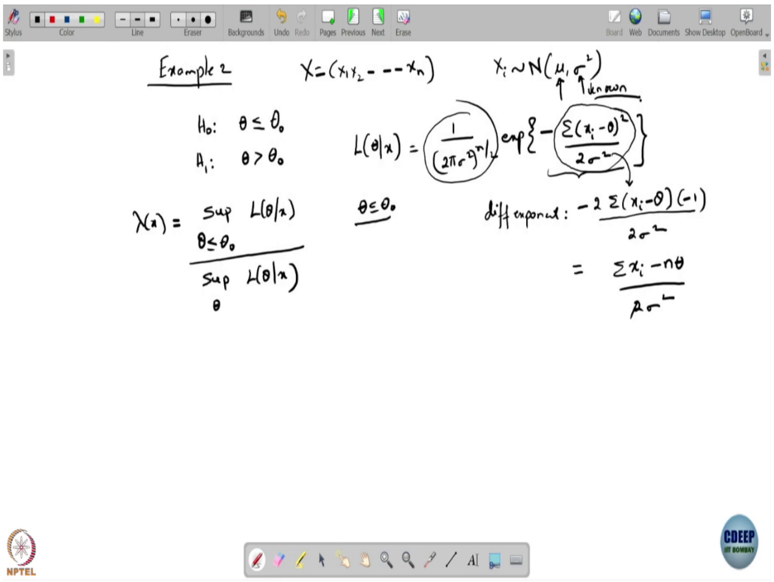
And our decision is going to be now if lambda x is going to be less than or equals to c, I am going to reject it to be null hypothesis and accept it as an alternate hypothesis. And notice that there is a c here which is the parameter that you are going to set for making this decision and what will be the range of lambda x here, obviously, denominator is going to be larger than the numerator my likelihood functions are positive. So, this ratio is going to be between 0 1. So, my c will be also for some range between c 0 1 we will take.

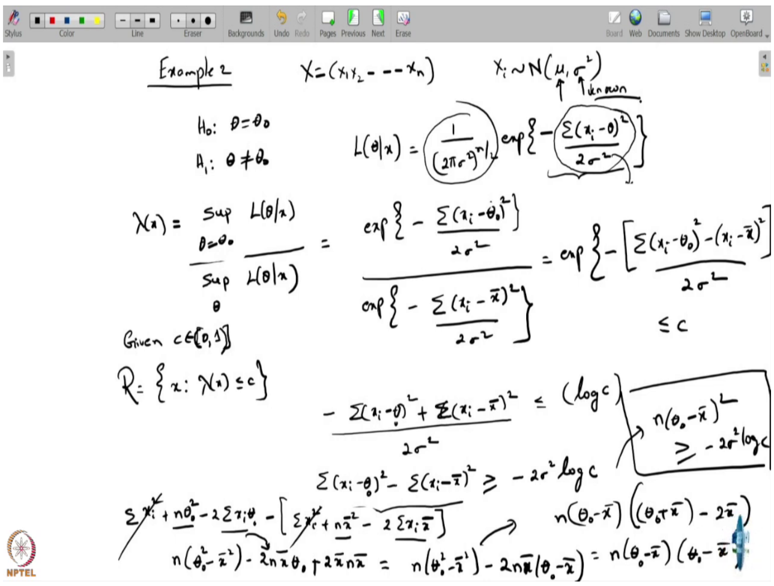
(Refer Slide Time: 03:39)



So we calculated last time this example about the binomial random variable. Sorry, we did it for a Bernoulli random variable with parameter p and our hypothesis were whether it is a fair coin or an unfair coin. So, what we will do is, today we will do one more exercise. On the…

(Refer Slide Time: 04:06)





Let us say your samples now, let us take and sample and assume that this xi's are coming from mu sigma square. And you want to ensure that, so, okay fine, your xi's are IID coming from gaussian distribution with parameter mu and sigma square and now your goal is to identify your hypothesis on this parameter mu and let us assume that the sigma square is known. And now I want to make two hypothesis, my null hypothesis is theta is going to be less than or equals to some theta not and my H1 is theta is going to be greater than this theta not, fine.

Now, first I want to know find out a likelihood ratio apply a likelihood ratio test on this and I want to find out that (likeli) what is that likelihood ratio and for that I need to first find out my likelihood function. So, what is the likelihood function of the theta given x, we know this is nothing but 2 pi sigma square n by 2 exponential minus xi minus, mu or theta here? Theta square divided by 2 sigma square. So, now let us write lambda of x this is sup over theta divided, theta less than or equals to theta not, this whole quantity and sup over entire theta L theta by x.

Now first let us do the bottom one, denominator one because that is unconstrained whereas the numerator is the constraint one. So, if you try to optimize our theta, what is the theta that will optimize this likelihood function? Sample mean. So, that is the one, the one that denominator is going to be a sample mean. What is the one, that is going to optimize the numerator? Why? Anybody can think why that is going to optimize a theta not? So, just think about this, whether this quantity here is increasing or decreasing and theta, decreasing? Let us say if you increase theta xi minus theta is going to decrease and with a minus sign increase.

So what is the effect of increasing theta on this quantity? Is it in increase? But notice that it is a square here it is not a linear term here. Does the square bother? Why do not I just work out quickly? What is that? Say this is a, the only parameter theta is coming here. I do not need to worry about this guy is a constant for me. Because sigma square is known. And now only thing you have to optimize is this quantity.

And, here, I am interested in maximizing but if you ignore this minus sign, you have to just minimize this quantity. Find out what is that theta that minimizes this, maybe let us quickly do that. So if I differentiate this, this is going to be simply xi minus theta times minus 1. I am just differentiating this quantity here everybody is in sync with me. Differentiating exponent I am just trying to differentiate the exponent, this is the exponent here or maybe let me keep that minus here. This minus I also kept, so 2 sigma square.

Now, so this quantity becomes what is become minus minus summation xi minus n theta not, sorry n theta, now, you have also the constant theta is less than or equal to theta not, how we are going to handle this constraint or 2 also get knocked out fine. So, how you do? How do you handle this theta less than or equal to theta constraint, you want to construct a langrage here, or not necessary?

Let us simplify this problem, let us say instead of this I will make your life easier, let us only consider this problem here theta is equals to theta naught and theta is not equals to theta not, is this simpler? I am only considering the hypothesis whether theta is theta not or not. So, then in the numerator there is nothing to optimize is actually a theta naught. And then what I will do maybe I will because of this simplification I will not go into all this circus.

So, again the numerator is going to be simply I will get rid of this constant because that anyway get counts canceled in the numerator and denominator. So, exponential minus xi minus theta not square 2 sigma square divided by exponential minus xi and what is that theta not that is maximizing? Sorry what is that theta that is maximizing this, that is exactly x bar, everybody agree with this? Now, if you simplify this is nothing but exponential minus xi minus theta not square minus xi minus xi bar whole squared divided by 2 sigma square.

This is my lambda, this is my likelihood ratio test. Now, I want to set my rejection region as or x such that lambda of x is less than or equals to c for some given c, I am going to set given c which is belonging to 0 1, maybe I should keep it open interval 0 and 1. So, now, let us write this what is going to happen?

So, now, if I want to now equate this to less than or equals to c, if I do this what is the condition I am going to get minus xi minus theta whole square plus the summation will retain xi minus x bar whole square and I will take the exponential other side this is going to be log c, and there is also that 2 sigma square factor is there, but I think I should not take it like this I will just keep it like this, is this side inequality fine.

Now, what I will do is I will take this minus sign with 2 sigma square on the other side. So, you will end up xi minus why this is supposed to be theta not, this is a fixed theta not, theta not minus xi minus x bar whole square now, this inequality become minus greater than or equal to, minus 2 sigma square log. Now, is there any simpler expression for this?

Student: Audio not clear.

Professor: Say, that again xi…

Student: x bar minus theta.

Professor: **x** bar minus theta not whole square…

Student: Audio not clear.

Professor: Oh, that is it, this will just happen to be x bar minus theta not. Do you agree that it is going to be simply x bar minus theta not this left-hand side?

Students: Audio not clear.

Professor: Why do not you quickly work out? I also want to work out this. Let us do, so, the left-hand side is going to be xi square summation plus n theta 0 square plus minus 2 summation xi theta, the first term. And the second term is going to be summation xi square plus n x bar square then minus 2 xi x bar, this whole thing and now this guy gets knocked up with this and I will club these two guys together, n theta 0 squared minus x bar square.

And now I am going to these two guys, so these two guys how I can club. So, 2, I am going to divide this guy by n this sum because theta 0 is a constant, I will divide it by n. So I am going to get x bar theta naught this guy to this and this guy, here x bar is also constant for me, from the data. So, that is going to be minus 2 x bar n x bar, you are right, this should be plus. And this you simplify this is going to be n theta 0 square x bar square minus 2 n x bar, 2 n x bar I have taken and theta not minus x bar, is this correct? That is it, now we have this it is not as simple as, as you said that it is simply theta…

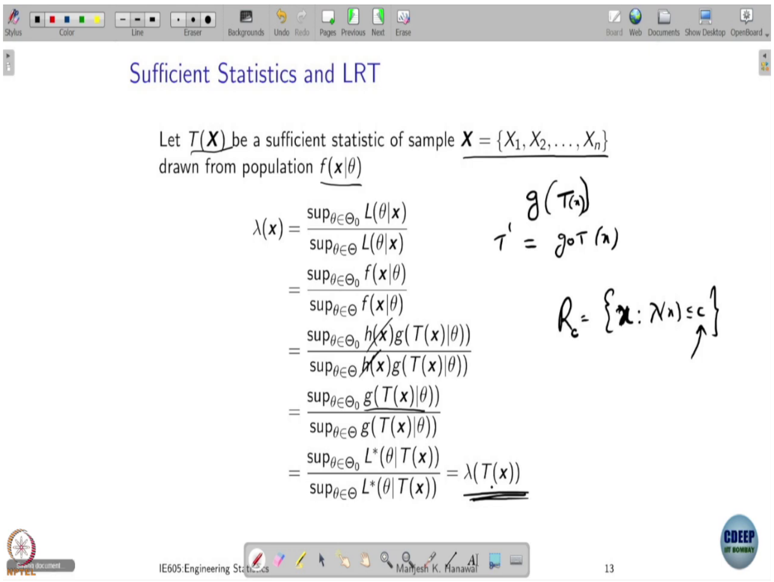
Students: Audio not clear.

Professor: You want to write it as theta minus x into theta plus x, I mean product of theta minus x into the theta plus this, let us do that. So, I am going to take theta x bar outside and n is anyway common this is going to be theta plus x bar and minus, n is already out to x bar. So, fine this is going to be n, is it coming to that guy, theta not minus x bar is that correct? So, you are his friend want to prove him right, fine, then this is going to be n theta 0 minus x bar whole square,

And this, so, final condition is theta 0 minus x bar whole this should be greater than or equals to minus 2 sigma square log c, this is my condition. If you want to further simplify, you can take the square root both sides divide take this n on the other side and you will get everything just in terms of if x bar is greater than something, but this is the condition you are going to get, if your x bar is such that this condition happens.

Then you are going to reject it that, that is equals theta equals to theta 0. And if this condition is not true, then you are going to accept that that is indeed has a parameter theta equals to theta not, fine. Any questions on this, now, let us look into.

(Refer Slide Time: 18:57)



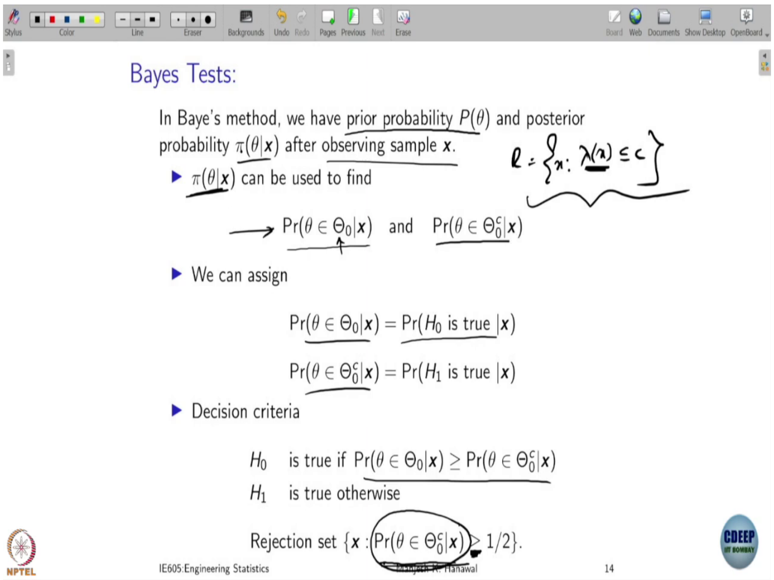
If I want to start with the likelihood ratio, sorry if I have if I already have a sufficient statistic and if I want to use it, what is the connection between that and my likelihood ratio test? So, let us say T of x is a sufficient statistics on my sample x, which is drawn from a population f x given theta. Now, our likelihood ratio test is this numerator and denominator. And by definition, this L of theta x we are simply taking it to be the PDF. But now, we are talking about having a sufficient statistics, L sorry T, and we if you recall one of the properties of sufficient statistics is that we can split my PDF into two parts.

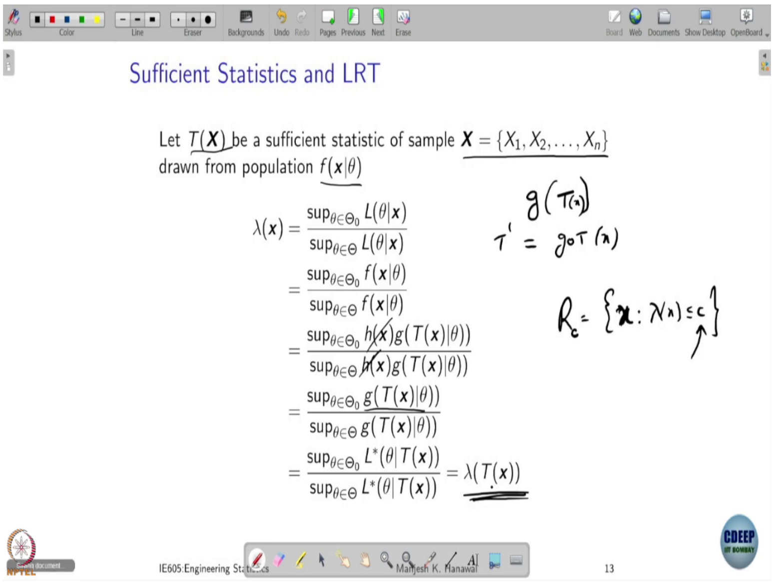
h and g, where h is a function of x alone, and a g is a function of x only through T. So I can write this in that case, I can quickly knock off this h of x because it has nothing to do with my parameter theta. Now, what I have is g of T of x, if d is a, T is a sufficient statistic, and g is another let us assume let us assume hypothetically g is a one to one function, what is this case g of T is, so, we have said like if T is a sufficient statistics and we are going to apply another function on this, this g composition T of x, this is let us call, this is T prime, this T prime is again a sufficient statistic.

So, now we have a ratio in terms of two sufficient statistics, you can calculate I am simply writing it as L star now, I have written it as L star. Maybe I should have L star now, this is a function of this given sufficient statistic. And now I am going to call got a new expression for my ratio test in terms of the sufficient statistics. That is what I am calling it as L of T of X. So, I am this is just I am denoting, like when I am involving sufficient statistics T in my computation, this likelihood ratio is coming from this sufficient statistic T.

Now you can do all the things you want, like you, you just define your boundary rejection ratio, always like maybe we should write the rejection ratio with a subscript c because that depends on that parameter c, if you change your this is all x all x such that lambda x is going to be less than or equals to c, if you are going to change your c, your rejection region can potentially change.

(Refer Slide Time: 22:19)





This is one way, now, this is kind of assuming a frequentist approach like assuming that. Theta is something some fixed parameter, which I do not know and that is the parameter which is governing my PDF. But we have always seen that instead of assuming theta to be fixed, we could be assuming that that is coming from another distribution itself. And in that case, we use this Bayes approach. So the Bayes approach could be also used in the hypothesis testing. How to use it, let us see.

So, in the Bayes hypothesis testing what we do, we are going to assume some prior probability on my parameter theta, that is always we do and after like you observed some sample x maybe your posterior is going to be pi of theta given x. So that we have computed, how to compute the posterior given my prior information and an observation x. Now, I can now define my new hypothesis, instead of exactly whether it is except it belongs to this or this, instead of asking this binary decisions, we can make it probabilistic decision now, weather tell me the probability that the observation I am making is coming from this parameter space.

Earlier it was yes no, but now we are putting some probability on the yes and some probability on the no. So that is why we are saying now, once they have this posterior probability after observing data or your sample, you can ask the question, what is the probability that theta belongs to my null hypothesis given my observation x, is can you compute this probability? From your posterior probability? So once you know given x, this posterior probability is giving you probability distribution of this theta.

Now all I am asking is that it belongs to the set theta 0. I should be able to compute this by integrating this probability or my space of theta 0. And similarly, I should be also able to answer the question what is the probability that theta belongs to my complementarities, complementary set using my posterior probability. Now, we can assign instead of asking whether H0 is true or not I can go in to ask the question, what is the probability that H0 is true given x? What is the probability that H0 is false here. Sorry, this is correct.

This is H0 and the what is the probability that H1 is true given x. Now, these are probabilistic arguments. Now, the decision criteria has to come earlier my decision criteria was I was putting some threshold or my likelihood, sorry LIT function log, my earlier my decision criteria was like this, where I could probably compute my lambda x, but here I am computing the probabilities given an observation x, what is the probability that H0 is true and H1 is true.

Now, what would be what would be a natural candidate to make a decision here whether H not is true or H1 is true? You just may want to go like if this probability happens to be higher than this, then you will say that, H0 is true, otherwise, H1 is true. But then that so, you will simply say that H0 is true if theta is belonging to theta naught given x is going to be larger than theta belonging to is complement set given x. And otherwise H1 is true.

Alternatively, you can just say that, if I have to put it in this structure, I will compute this probability. And if this probability is greater than half, then I am going to reject it or basically accept the H1 hypothesis. So, now, this lambda x is being replaced by the probability that my parameters are coming from a null hypothesis given my observation x. So, in the frequencies approach, we will just construct my likelihood ratio test.

And in this Bayesian approach, I am just going to compute this probability based on my posterior distributions. So this part on the Bayesian test, any questions on this? You can make it strictly greater than, if you are going to set it like this. Maybe you can just make it strictly greater than, that does not matter. Usually we were going to be like, that matters if your theta space is discreet, but most other times we will be working with theta space which is continuous in that case, it does not matter.