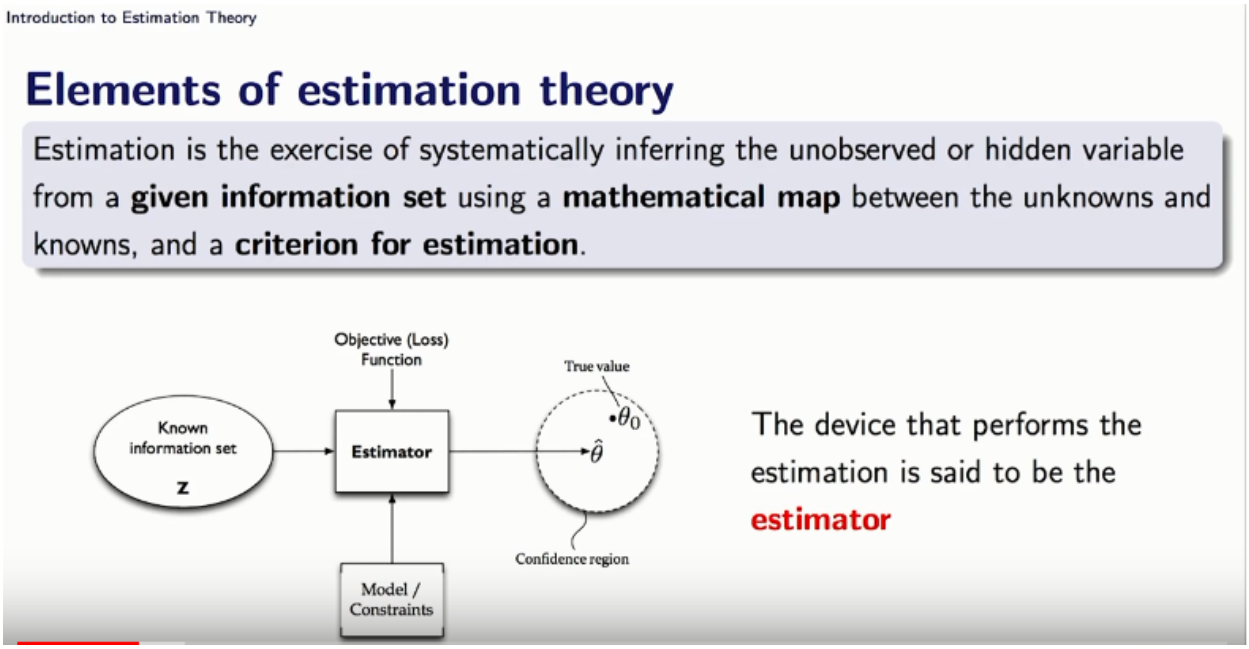
Lecture-37
Part 1 – Introduction to estimation theory 1

Okay! Good morning yesterday we had a very brief introduction to estimation and what estimation theory offers. what? we will do today is we will first work out a simple example that hopefully will still they idea of estimation how a typical estimation problem is formulated. If you recall yesterday, I have said estimation problem is an optimization problem. A lot of estimation problems are formulated as explicit optimization problem. But there are some which are not even though they are not explicitly formulated as an optimization problems, and the heart of it and there is some optimization that's happening. Right? So, the example that we are going to look at will through like on several aspects of estimation, it's a very simple example which really like because of its simplicity and because of its ability to tell us what is involved in estimation and what are difficult different elements of estimation.

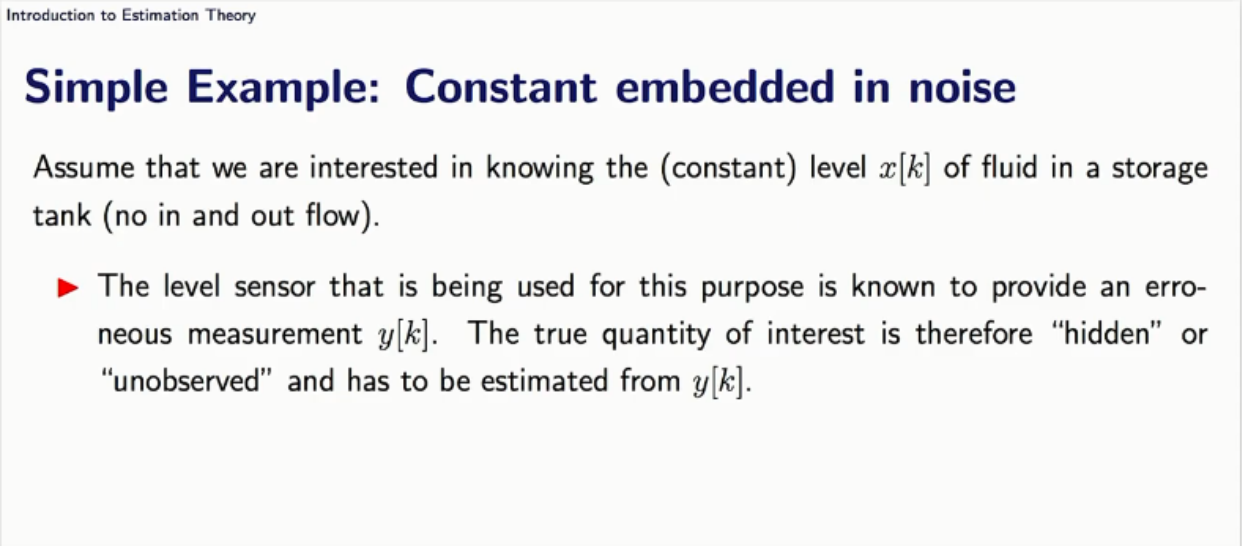
We have talked about five elements of estimation. Right? We said there is data, which is a food for estimation, and then you have the model an on their constrains which basically constructs a bridge between the space of data, the space of knowns to the space of unknowns and then you have an criterion of estimation or criterion of fit whatever you want to call an objective function, and then put together everything is bond the estimator which is like a formula ultimately your estimator is a mathematical expression it's a formula in probability you must have had the term statistic, a statistic is also a mathematical function which takes in the sample that is observations and then produces a number but the difference between the statistic and an estimator is that a statistic is simply a mathematical function which takes in the data and produces an number there may not be an purpose to it, whereas with the estimator there is a purpose and that purpose is to arrive at an intelligent estimate of the parameter. That's a certain difference otherwise there is no mathematically any difference between the statistic and an estimator. And once you have arrived at estimator then you plug in the data and then out comes the unknown estimate of the unknown which we call as the point estimate. We also said that there are internal estimators, so there the estimators were you plug in the data and it will not give you a single value but a range of possible values.

Video Start Time: (3:16)



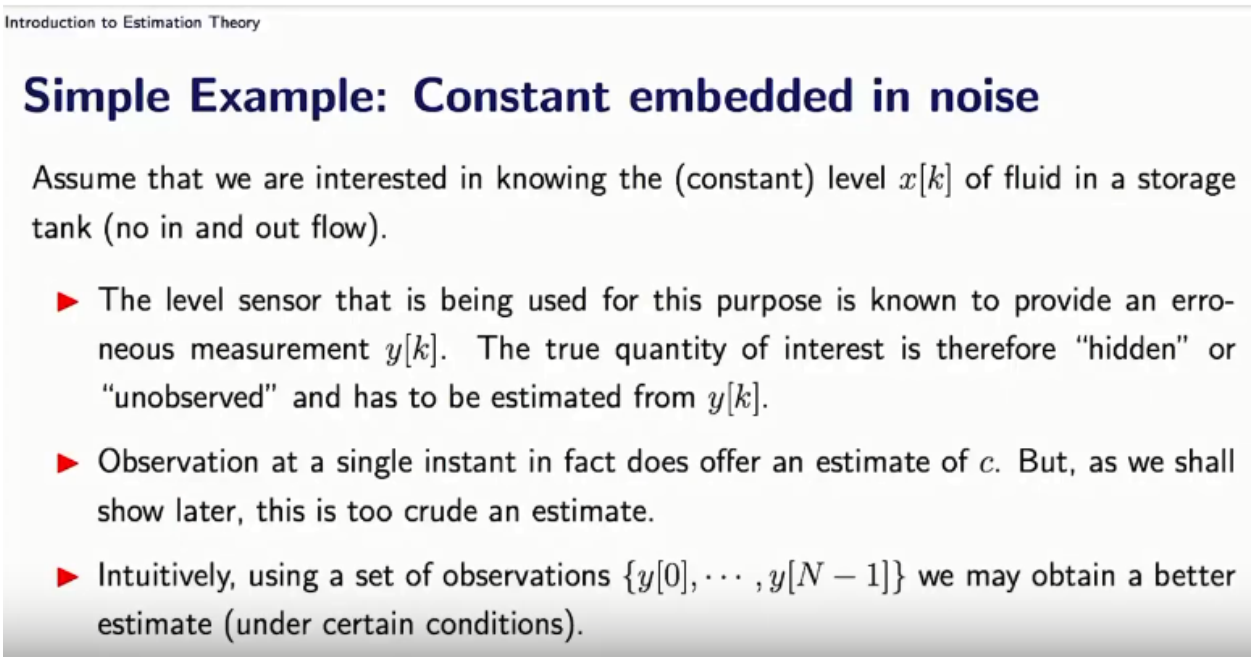
We will first look at the point estimator and then understand what happens later on. So, I am going to skip pass these slides which describe each of this elements I have given you already the essence of this and therefore its best to move on to the example. There are some aspects that I will highlight after I go through the examples.

Video End Time (3:46)



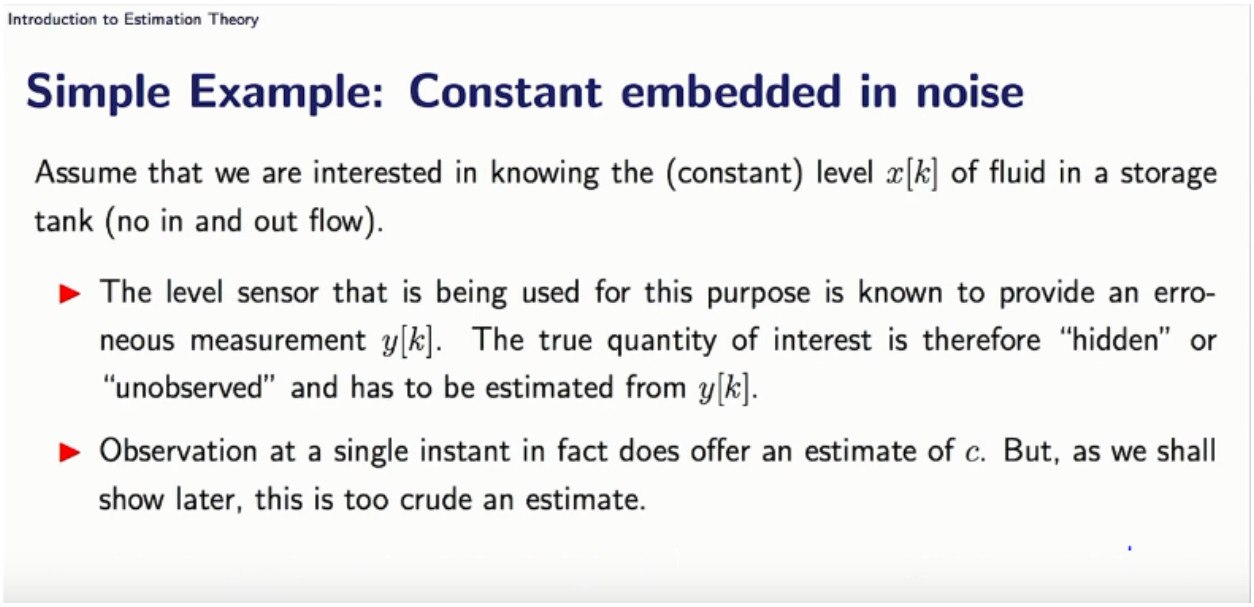
So, let's begin with this example of a constant signal embedded in noise. As I said yesterday this example is actually very wonderful because it's a I am writing this upfront but actually it comes at a later stage, more descriptively what you have is there is a "constant signal embedded in noise what kind of noise we will not know". Right? at that moment I have written a model here but let's not jump to that. The problem is that of now estimating this signal constant signal this could be of level reading, temperature any physical variable does not matter, its embedded in noise. So, for a signal processing person this is signal estimation problem. Right? But we can turn around turn the problem around and say for a statistician this is the problem of estimating mean of an random signal. Right? Momentarily if you assume the model that I have written there for the measurement since ek is typically assume to be zero mean and goal is to estimate see you can think of the problem as also estimating the mean. So, and then for a state space person working in state space model you can say look I have a measurement y and state is c. Right? Typically, when you think of state space model you have state equation as well which tells you the dynamics of a state. But this is a special case were the signal is at the steady state. So, there is a state, but that state is at steady condition, and that is the see that you want to estimate. So it has all the flavors it is a signal estimation problems if you want to look at it at that way it is a parameter estimation problem if you like it that way it's a state estimation problem if you want to look at that way. So, it's a very beautiful example and it is not a complicated one. Right?

Refer Slide Time: (5:54)



Simple examples are usually very helpful now I just to put this signal sorry,

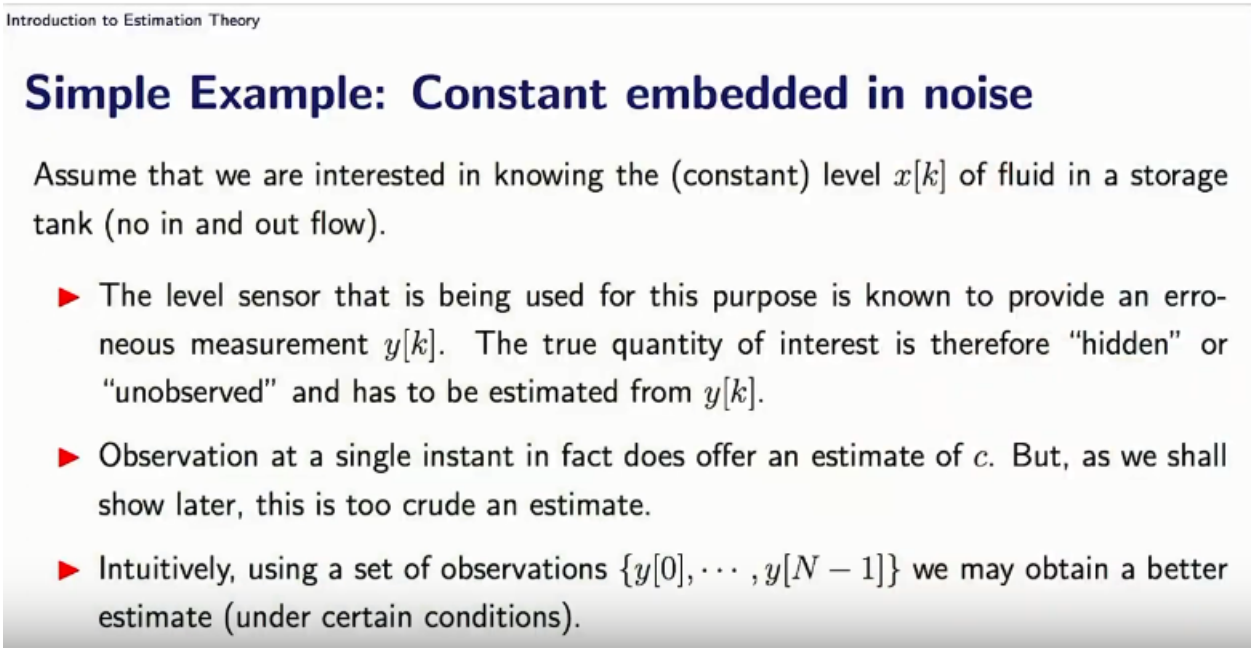
Refer Slide Time: (6:00)



Just put the example in context I would like you to think of this measurement as set of a level measurement, you know there is a liquid level you don't have to put little context but it helps so that you will get a feel of the problem. I have a liquid level system and the level is fixed, there is no inflow and there is no outflow will neglect evaporation losses, so we know there is no vibration and anything like that. So, the level is fixed, and you are sensing it. What? You are sensing is y what you want to know is c.

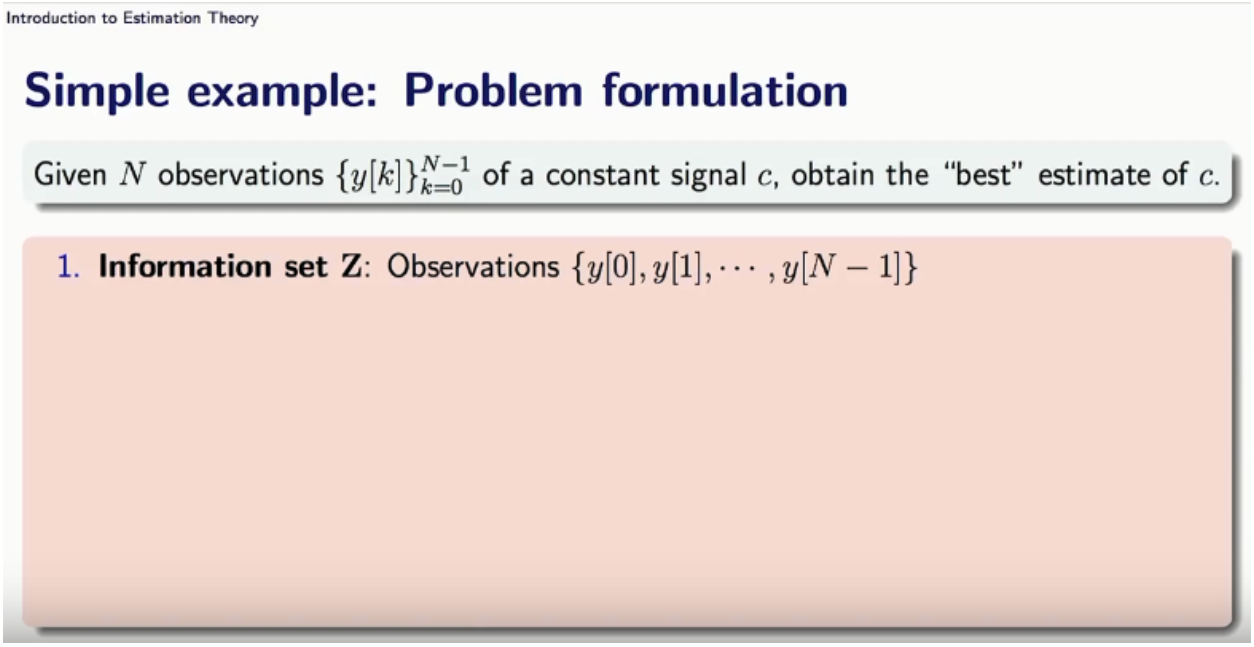
Okay? Now one esti, one crude estimate of c very crude estimate if I am very lazy is, I will pick randomly any of the measurements and say that is the estimate of c. Right? After all that is what we do in life whatever reading I get for the thermometer for the body temperature from the thermometer whatever reading I see I say that is the estimated body temperature. Am I actually collecting when I have fever, I am I do I will I have the patience of actually collecting some 100 observations wait, wait? Right? And the sister will never tell you wait I have to collect some n observations and I am going to take the sample mean that's not going to work. Right? In such situation you will just take a single reading and call it as a estimate. How? crude is that estimate we will learn later on. At the moment I am saying it's crude estimate of c, crude in what sense. Right? We will understand what it means but nevertheless it's a fairly good estimate to begin with. We are not going to do that we have lot of time it's a we, we will we will take all the observations and come up with

Refer Slide Time: (7:58)



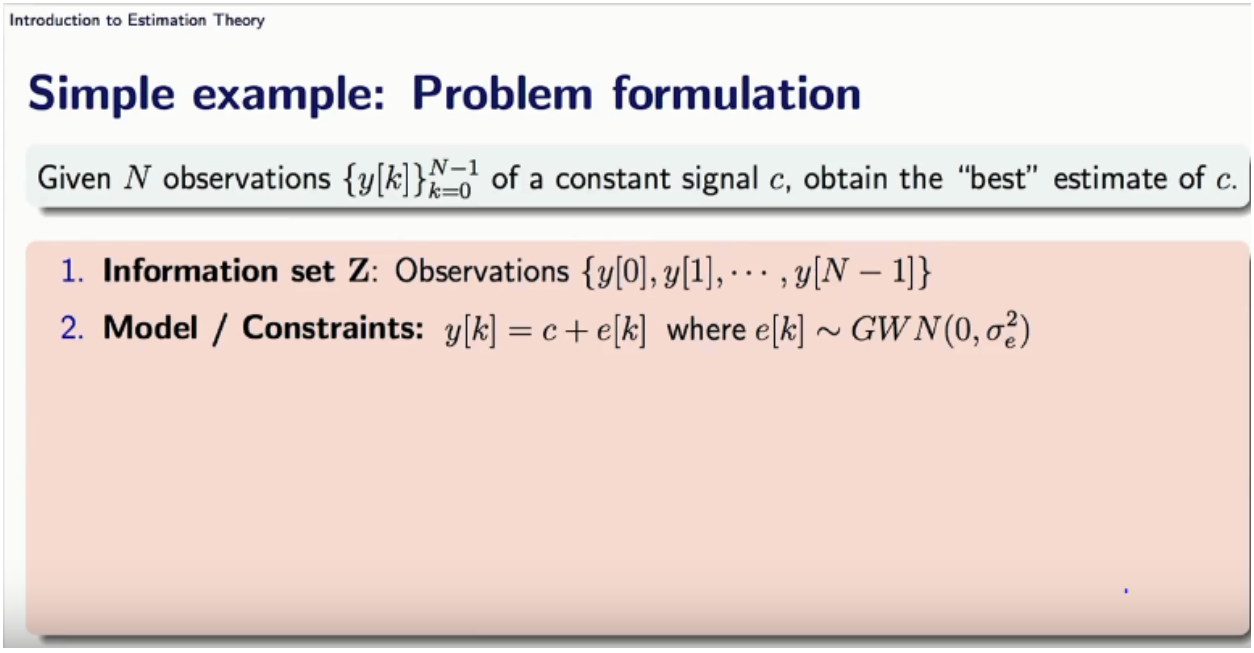
A better estimate hopefully better in what sense we have to argue everything in estimation is stated very clearly. If I say better in what sense if I say good in what sense. Right? If I am talking of precision, I have definition if I have an accuracy if I say the estimate is accurate, I have the definition. So, everything is been clearly articulated and that what we need to know now how to articulate the things in estimation theory using definition.

Refer Slide Time: (8:31)



Okay? So, the information that has been given to me if now I break this up into elements of estimation what is the information that I have data, n observations. Right? That's the first element.

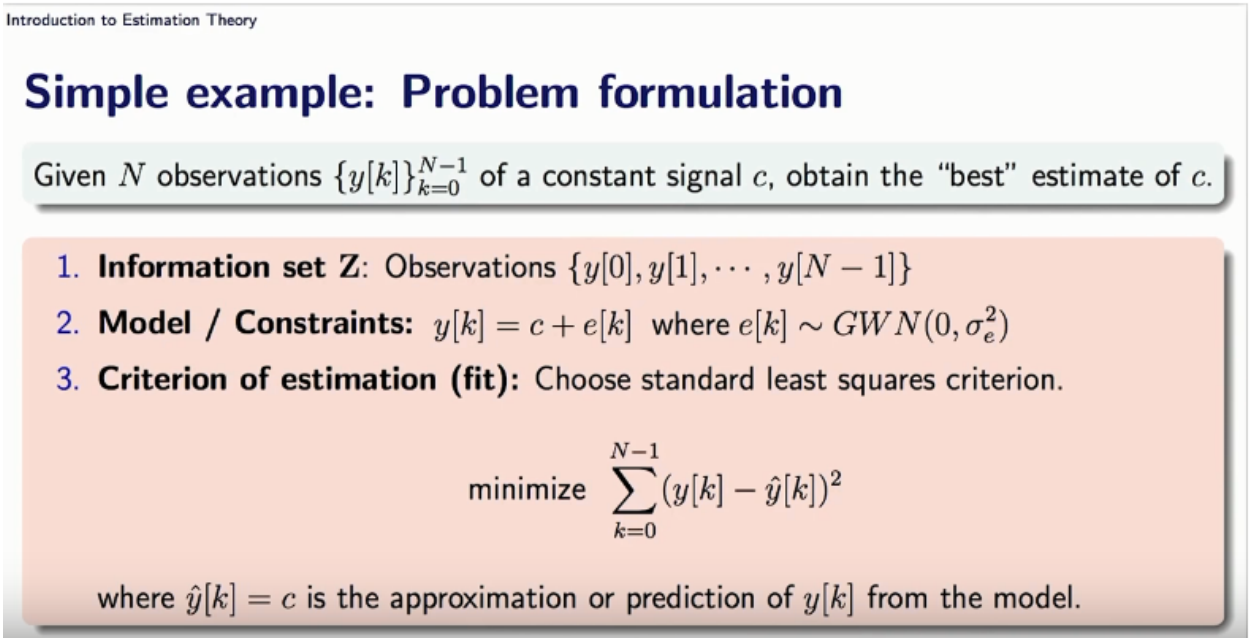
Refer Slide Time: (8:47)



Now I have to assume some model some generating model for the data. The purpose is now to connect the known space which is data to the unknown space the parameter that I want to estimate what do I want to estimate the there is a constant signal underneath these measurements, and I want to estimate. Now this is the model that we assume which you see on the screen. What is that we are saying we are saying first of

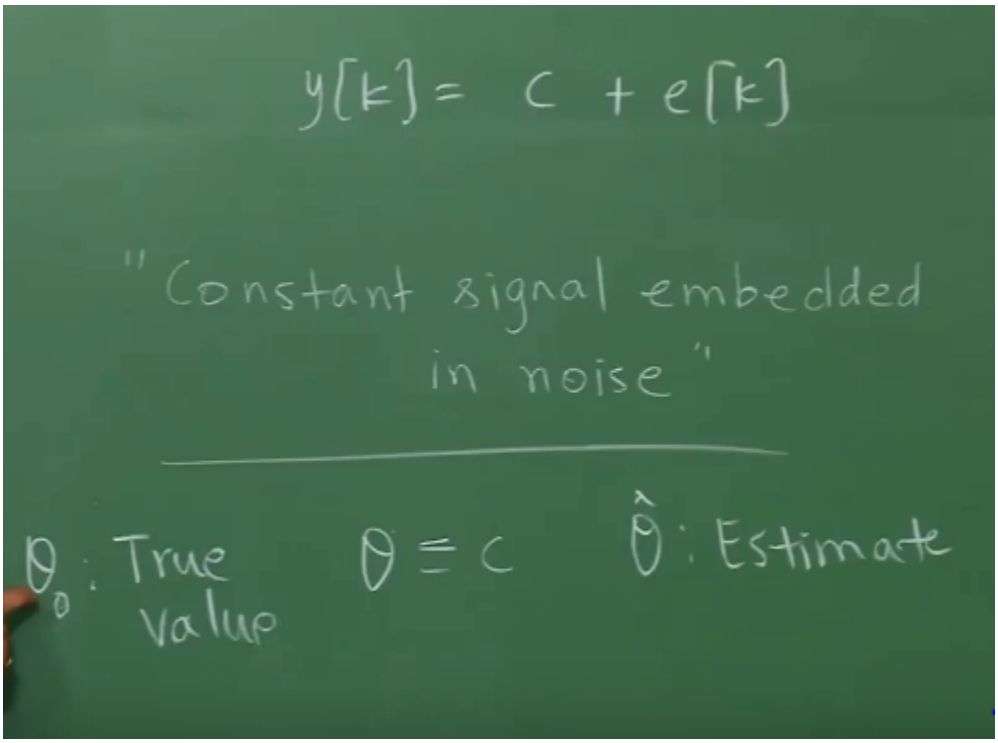
all that the noise acts on to the constant signal that doesn't multiply? So, I am assume an additive noise. Right? You could assume multiplicative noise too and nothing prevents you from doing that, so which means now you start to see the branches. Right? There is a huge tree on which several branches are there. So, I am sitting on the branch of additive noise. I am going to sit spend my time there what else have I assumed by writing this what else have I assumed correct. The noise is white, but is that the only possibility what do you think? Is the white noise is the only possibility I could have colored noise too? But always academicians have this knack of assuming the simple things in a class and throwing the complicated one's during the exam, I am not scaring you or anything but or giving you the homework.no the point is we want to keep things simple so that we first understand what is estimation all about. So, we will assume white noise and, now in your mind, slowly you have to tell yourself look there are several possibilities of which I am assuming one. If I change my model its very lightly that the final estimate that I get is going to change. Okay? But let's stick to this

Refer Slide Time: (10:57)



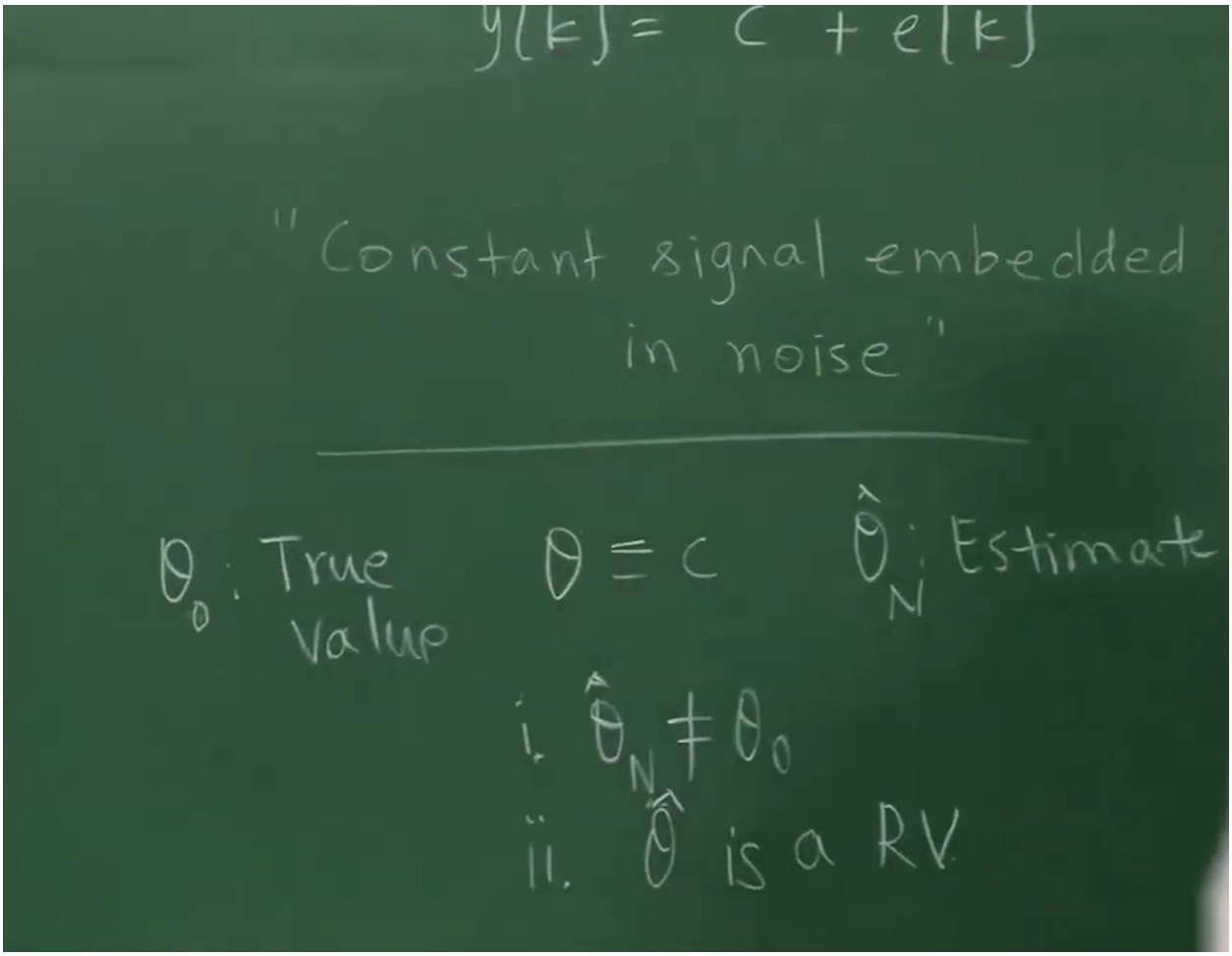
Then next we have the criterion of it. Now here is were you have to really understand an important thing. Right? Straight away I have written the criterion but let me spend a couple of minutes in explaining to you.

Refer Slide Time: (11:16)



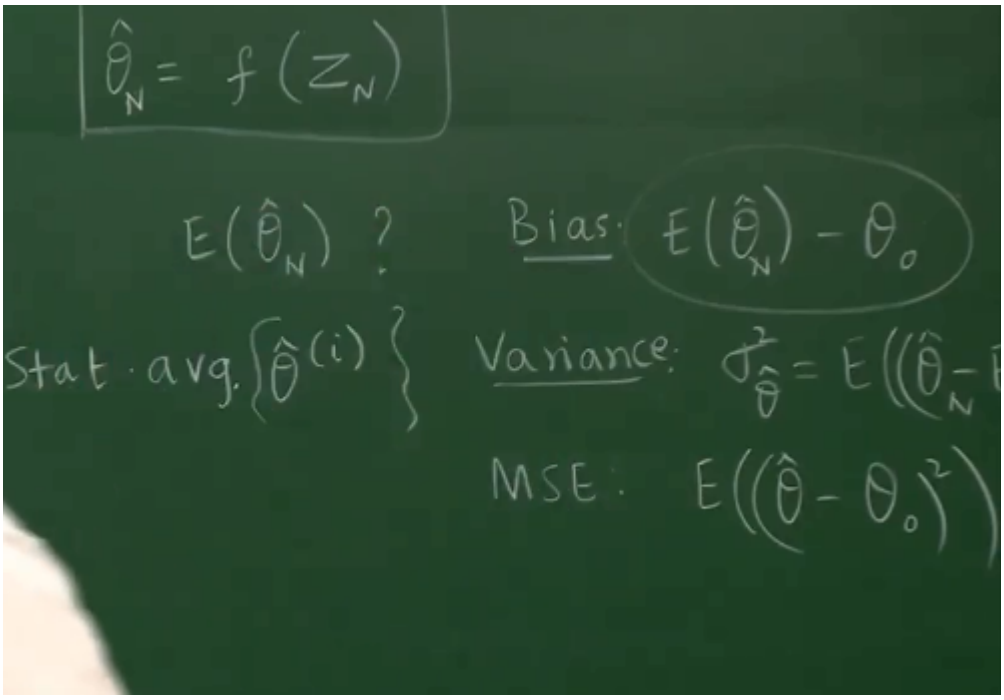
Theta is the parameter this is the notation that will use this is the parameter that we want to estimate which in our case is c, so this is the single parameter. Now let's call theta hat has the estimate the single number that I am going obtain eventually as a guess of c, ideally and n will also introduce this notion this extremely important. we will call $theta_0$ as true value. I do not know the true value but there exists some true value at least in simulation I know what the truth is correct. Ideally, I would like to obtain theta hat such that it is $theta_0$. Right? Is in that the ideal requirement but idealities are never realized there is no ideal gas, there is no ideally linear system, there is no ideal husband, there is no ideal wife. Okay? All of us have our own imperfections, systems also have imperfections, but it is good to begin with ideality so that we know how far away from it. So, one thing is for sure theta hat will never equal $theta_0$ when you are estimating it from finite observations. Why is that? You have ask yourself why is that I will never get theta hat as $theta_0$ or under what conditions can I get? When there is no randomness then I can afford to think of getting the truth. When there is randomness, I am going to work with a small finite length realization of the randomness, and that will never fetch me the truth. Right? So the first fact of estimation is theta hat subscript n sometimes this is indicated with subscript n to indicate the there is been observed from n observations theta hat n so these are certain axioms or you can say facts the theta hat n will never equal to $theta_0$ for finite n as n gets to infinity there is possibility will talk about that later.

Refer Slide Time: (13:46)



Secondly which is extremely important for you to understand and it is the central thing that you know rules? all these statistical inferencing hypotheses stressing and so on which is that theta hat is a random variable. Why theta hat is random variable? On what basis have I made this statement any idea. Okay? Remember now you have to ask how is, theta hat where is theta hat popping out from, theta hat is popping out from data. Correct?

Refer Slide Time: (14:33)



It is function of data ultimately it could be a simple linear function, nonlinear function we don't know ultimately I am going to use a formula, f for formula, f function that function could be linear nonlinear whatever but the feed to that function is data, which is going to have a randomness and we know already that any function of random variable is also a random variable. Correct. It will inherit the DNA therefore theta hat is a random variable what about theta_0 in the classical estimation theory theta_0 is a deterministic quantity its fixed for a given process its fixed in a bastion estimation theta_0 also considered as a random variable in the bastion world but we will talk about it later let's not bother ourselves about it right now.

For now, theta_0 is fixed it's the truth and I am trying to reach that. Okay? So, you should always remember theta hat is always a function of the data it could be linear nonlinear whatever, but it is a function and therefore it is a random variable, which means now theta hat will have its own mean, its own pdf. Right? Its own variance everything that important thing is to learn how to interpret what is the mean of theta hat what is variance of theta hat. So, let me ask you what you understand by expectation of theta hat? So, we are very soon going to learn expectation of theta hat what is this? Will soon drop the subscript n for now we will continue to maintain. What do you understand by that yeah you can say it is a random variable it's a fast moment of pdf it is center of outcome and so on? But how do you explain? what you mean by expectation of theta hat any idea? Sorry? you want to try how do you explain how do you imagine as a thought process do you agree that? theta hat is a random variable it is it has its own mean how do you interpret it now? why is theta hat is random? because the data has randomness what do you mean by data

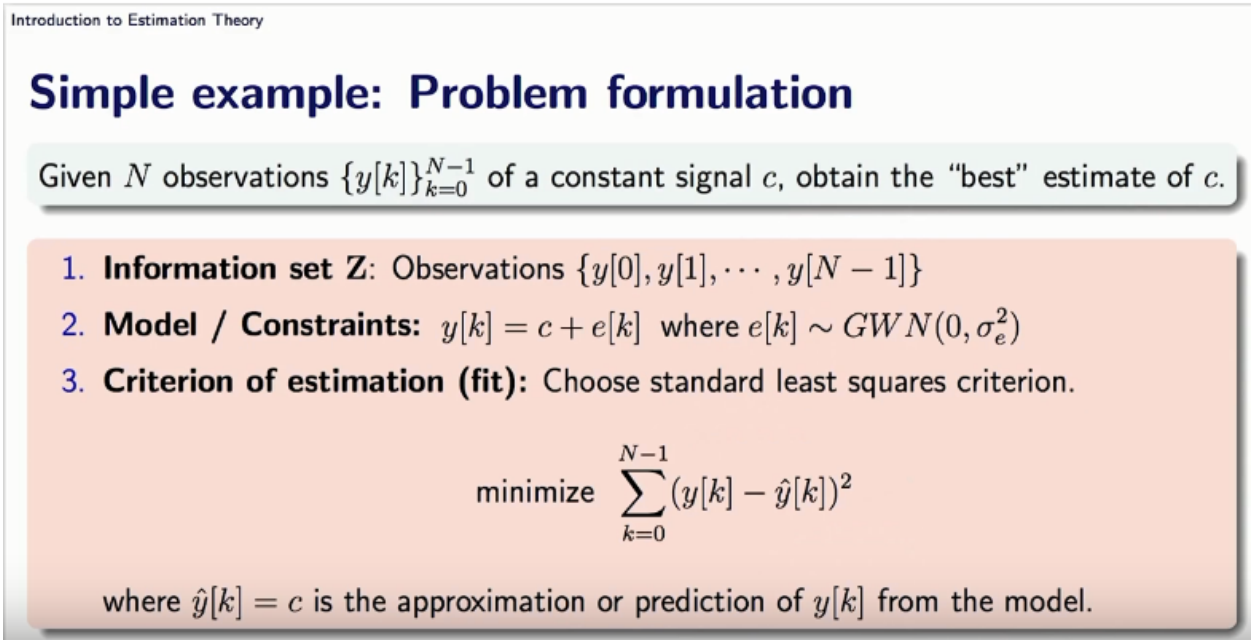
being random? what do you mean by data being random, do we understand what is mean by data being random. How do you explain it say data is random you should I am sorry?

Correct. So yes! so you have only one of the many possible records that means if I have to repeat the experiments with everything, will I have fixed I will obtain a define record, that what is the trademark of the random process. Right? So, now using that you should be able to explain what the expectation of the theta hat. if I have one record I have one theta hat if I repeat the experiment, I have another data set I will have another theta hat. Correct. Another data record another theta hat so for every realization of the data record I have different theta hat expectation of theta hat is the statistical average of all those theta hats. So as a thought process you should get used to this thought experiment the thought experiment is that I repeat the experiment I get one data record I calculate theta hat repeat the experiment calculate theta hat. So, for every i^{th} experiment, So, we have for the i^{th} experiment the theta hat i this expectation is the statistical average of this theta i's were do we have the noise I don't know this noise allocates the amplifier. Okay? So, we in principle we may have infinite possible values of theta hat. Right? We don't worry about that as long as we understand that its okay the same way you can understand variance of theta hat. If theta hat is the single theta is the single parameter, we can say variance of theta hat. If theta is a vector of parameters, then we talk of covariance. What ever may be the case how do you explain again the same story I repeat the experiment I have theta hat another theta hat the variance of theta hat tells me how is the theta hat varies across the experiments. Is in that important? What do I want to do I want the variance of theta hat to be high or low, low because if it is very high that means I cannot rely on a estimate obtain from a single record? If I had to repeat, then it should be completely different value. So, we want the variance of theta hat to be as low as possible. Okay?

So later on, you encounter a terms such as bias which would be the difference between the average of theta hat and the truth and then you would have variance which would be sigma square theta hat again like a in a random variable this is nothing but the same story expectation of theta hat minus mu theta hat to the whole square same there is nothing these expressions or not any new the moment you recognize theta hat to be random variable you are set. You will also encount a term such as mean square error which is expectation that is the spread of theta hat with reference to the truth. So, what is the difference between variance and MSE mean square error? Is that a difference in the definition? what is the difference? in variance I am using its own average as a reference, in mean square error I am using truth as a reference. But what are both measuring the spread of theta hat. That means how far the estimates are spread we don't want them to be spread to far why? because I with great difficulty I perform an experiment and I want the estimate obtain from the single experiment to be reliable. If it, if there is high variability, then the reliability of the single estimate is going to be low I can't trust this estimate to much this is what

technically known as precision. Right? This is got to do with precision variance on MSE whereas as bias is got to do with accuracy. Earlier we said that theta hat equals cannot equal theta_0 for obvious reasons, But we can expect this quantity here to be 0. I can expect that, that means if I have to average theta hat across all data records, I can expect at least to recover the truth. If I am able to do that, we call such estimator as unbiased estimator. Theta hat will be in error with respect to theta_0 because of randomness but the average of theta hat if it is different from theta_0 that means there is a systematic error not just a random error. Okay? So bias is looking at systematic errors we will talk more about it lets return to the example here. Right?

Refer Slide Time: (24:07)



Now, Why I have to give you so much preface is because ideally I would like to estimate c such that this is minimized ideally I want to c such that or you can c hat doesn't matter but will not use c hat here whatever I obtain as a solution will call a c hat. So, ideally I want an estimate in such a way that it is at a minimum distance from a truth. But do I know the truth? I do not know, so as it stands it as the problem is ill post. Correct? If I know the truth will I estimate there is no estimation problem at all. So, I don't know the truth and therefore this problem is unsolvable at the moment later on we will find a solution in a different way. If I want to solve that problem, we will realize that we may have to lets c_0 also being a random variable. Then it may be possible to solve it, that is what the idea behind the bastion approach. So now what is the way out you say yes! I agree that I do not c_0, therefore I cannot solve this problem, but I may know something about c_0 that means there is representative of c_0 available to me. What is the representative available data, so I am assuming data is to be generated by c_0 correct? So, I have yk which

is some function of c_0 that is in this generating equation if I am plug in c_0 I will be getting y. That is how I simulated the data. Right?

So, I don't have c_0, but I have yk, now we can pores a problem our estimation problem in a different way. So, if I know c_0 I know yk but in any case, I know yk. Now if I know if I have an estimate of c what I am going to do with the estimate I am going to predict. Correct. So, what will be my prediction if I am given c what's the best prediction what's the best prediction c correct. It is the conditional expectation ideally; I should also say y hat k given c and past data. But this is the steady state problem I don't have to worry about that. Otherwise I would had written a one-step I had prediction. Correct. So, what this tells me if I given c I would predict in y in this way. So, y hat is a function of c or c hat you can say. Now what I will do is I will drive y hat very close to yk that is the idea. Okay? So, find c. Right? Such that the above is satisfy do you understand we are not solving a direct problem we are solving; we are estimating c indirectly. We are hoping that as y hat is being driven close to y, c hat will also be driven close to c_0 whether that actually happens or not we do not know but we are hoping that is how now this criterion of it that you see on the screen is born. Now I have only written qualitative statement on the board, as I said drive y hat very close to y, but what is very close I need a measure, I need a measure of what distance. Correct. How when you measure when the distance are available. He is trying to answer. Okay?

I don't understand that answer so how many measures of distance are available, at least name one or two Euclidian, Right? The standard Euclidian distance then what else non aton distance which is one known. Right? And so, on mahalanobis distance there are so many distance measures we will pick the squared Euclidian one. So now what we are saying is all that I written on the board drive y hat very close to y, but we don't have a single reading, we have n readings, so we want collectively drive y hat close to y. If I am giving importance to single observation more importance to single observation a few observations and less importance to others that's a different matter. But to begin with I want to give equal importance to all the observations. And I say collectively drive y hat close to y and there comes your least square objective functions. This is called a sample least squares because I am working with samples observations do you understand? If I give more importance to y hat some, some readings and less importance to others. Then I have a weighted least square. When will I do that? What is the right situation for doing that why would I to give more importance to certain observations and less, lesser to others?

Can you think of a scenario? Correct. So, its likely that certain observations are less reliable then the others? Here we have assumed ek to be white that always means it have fixed variance at all times. But suppose ek is so called heteroscedastic which means its variance keeps changing the time still white, but the variance can change the time. Because may be the sensor is heating up or the sensor is actually giving you different errors depending on the readings if you take for example a temperature sensor like a thermo

couple at the ends it will have more error at the midpoint. Correct. For every sensor there is a range if you push it to the extreme then it will give you more error, and then there is an operating range or which you can expect uniform error. So, when you have errors with varying characteristics then you can think of a weighted least square subjective function. At the moment we will keep things simple, so this is now the criterion of it.