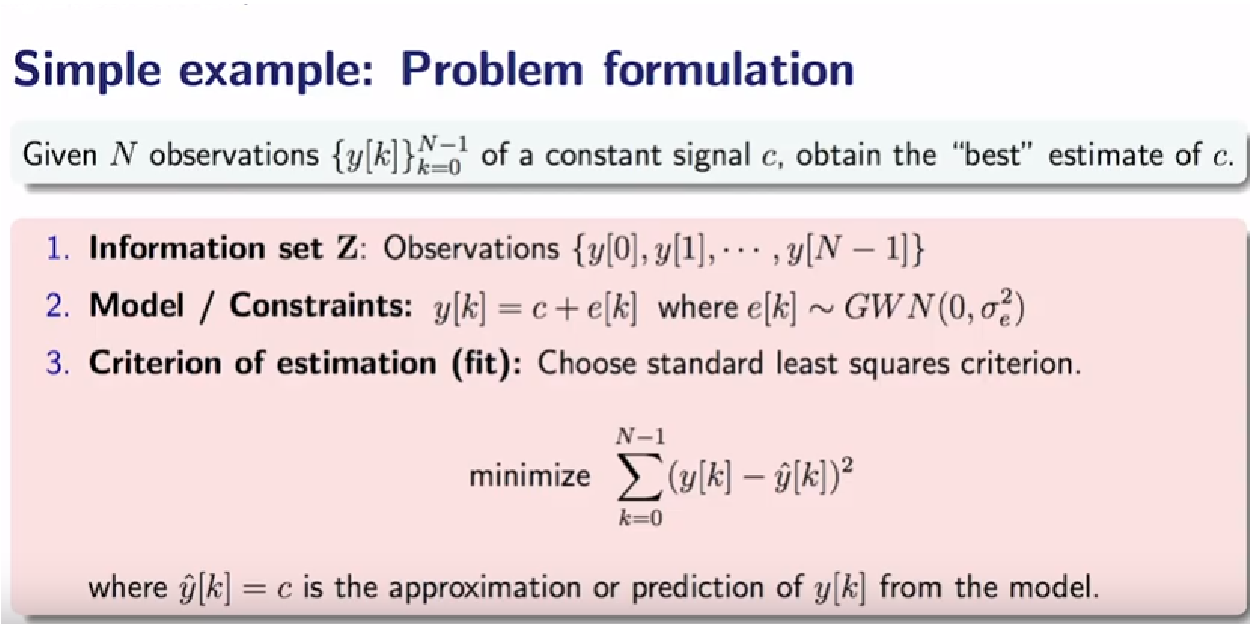
Lecture 37
Part 2 – Introduction to estimation theory 2

Now all that remains to be done is what? Solve this problem; can you give me the solution? What is y hat? It is C. so here is your objective function simple objective function, find the optimal value of c such that is objective function as cost function is minimised, its cost function is minimised. What is the solution? How do you solve it? How do you solve a standard optimization problem? Take the first derivative and set it to Zero, Right? Of course does it necessarily give me the minimum value of the function? I have to go past and check and look at the sign of the second derivative. But let us take the first derivative and give me the solution first. Very simple, in a minute you should be able to answer it. This is what you had to get use to gradually.

Refer Slide Time :(01:34)



I am sorry, correct. So all you have to do is minimize, take the derivative of this with respect to c and set it to zero. Y is independent of c, because Y is driven by c naught, correct? only Y hat is dependent on C, you can think C as your free parameter.

Refer Slide Time :(02:15)

Introduction to Estimation Theory

Simple example: Solution

Introduce

$$\varepsilon[k|\theta] = y[k] - \hat{y}[k|\theta] \qquad \text{(prediction error)}$$

$$\mathbf{y} = \begin{bmatrix} y[0] & \cdots & y[N-1] \end{bmatrix}^T; \qquad \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}[0] & \cdots & \hat{y}[N-1] \end{bmatrix}^T$$

Then, the **least-squares estimate** of $\hat{c}$ is the solution to

$$\min_{\theta} \sum_{k=0}^{N-1} \varepsilon^2[k|\theta] = \min_{\theta} ||\mathbf{y} - \hat{\mathbf{y}}||_2^2 \qquad \text{subject to } \hat{y}[k|\theta] = c \qquad (2)$$

So solution that you obtain, is of course you know can write this objective function and vector form and so on but that is pretty obvious,

Refer Slide Time :(02:21)

Introduction to Estimation Theory

Simple example: Solution ...contd.

Solution:

$$\boxed{\hat{c}^\star = \frac{1}{N}\sum_{k=0}^{N-1} y[k]} \qquad \text{(This is the sample mean!)} \qquad (3)$$

The function $\hat{c}(\mathbf{y})$ is said to be the **estimator**, while $\hat{c}$ is the **estimate**.

Let's move on to the solution, C star is one over N, sigma y (k), what is this? This is nothing but our sample mean. You could have said Oh! You know? I knew it intuitively it has to be the sample mean because after all it is one of the very popular estimators of mean but would you know that actually in what sense it is optimal? You give me sample mean as well as an estimator but will you be able to answer in what sense is best? here I am able to answer both, I have arrived at the sample mean in a systematic way I have not forced my thought on it, I have not said a priori sample mean has to be the solution, I simply said, let me pick a objective function, first let me pick a model, postulator model, pick an objective function and then the solution that happens to be the sample mean. So now I can safely say, Sample mean is the best estimator of the C, under this model and under this criterion of it. What does it mean? If i change the criterion of it for example, we talked about different measure of distance, Right? Suppose if I change the criterion of it to, I will talk about variants of EK estimate later on.

Refer Slide Time :(04:03)

Introduction to Estimation Theory

Impact of objective function

Changing the cost criterion has a strong influence on the final solution.

Suppose we choose to minimize the 1-norm instead of the 2-norm, while keeping the predictor fixed. Then, the optimization problem and the solution are

$$\min_{\theta} \sum_{k=0}^{N-1} |\varepsilon[k|\theta]| = \min_{\theta} ||\mathbf{y} - \hat{\mathbf{y}}||_1 \qquad \text{subject to } \hat{y}[k|\theta] = c \qquad (5)$$

Suppose if I change the objective function to one norm, see earlier we wrote sigma yk minus y hat of k square, it can be written has two morns of prediction, square, two morns of prediction error. This summation is in terms of every observation; the norm is in terms of a vector that is all. So if we change our criterion of it or distance from square two norm, one norm what happens?

Refer Slide Time :(04:39)

Introduction to Estimation Theory

Impact of objective function

Changing the cost criterion has a strong influence on the final solution.

Suppose we choose to minimize the 1-norm instead of the 2-norm, while keeping the predictor fixed. Then, the optimization problem and the solution are

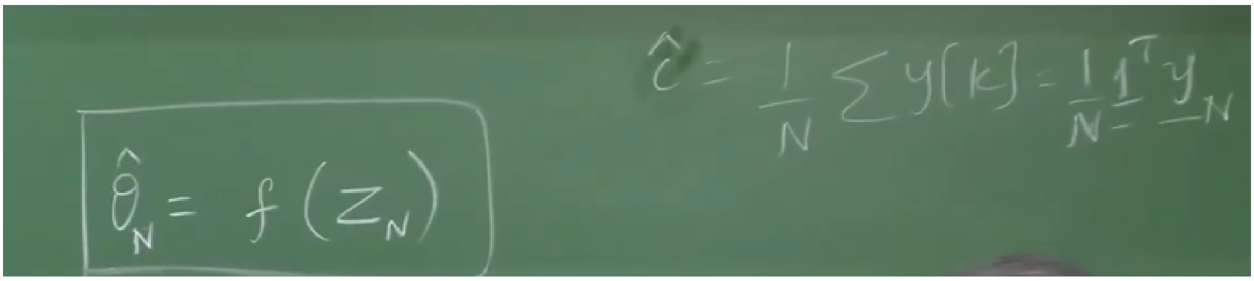
$$\min_{\theta} \sum_{k=0}^{N-1} |\varepsilon[k|\theta]| = \min_{\theta} ||\mathbf{y} - \hat{\mathbf{y}}||_1 \qquad \text{subject to } \hat{y}[k|\theta] = c \qquad (5)$$

Solution:

$$\hat{c}^\star = \text{Median}(\mathbf{y}) \qquad (6)$$

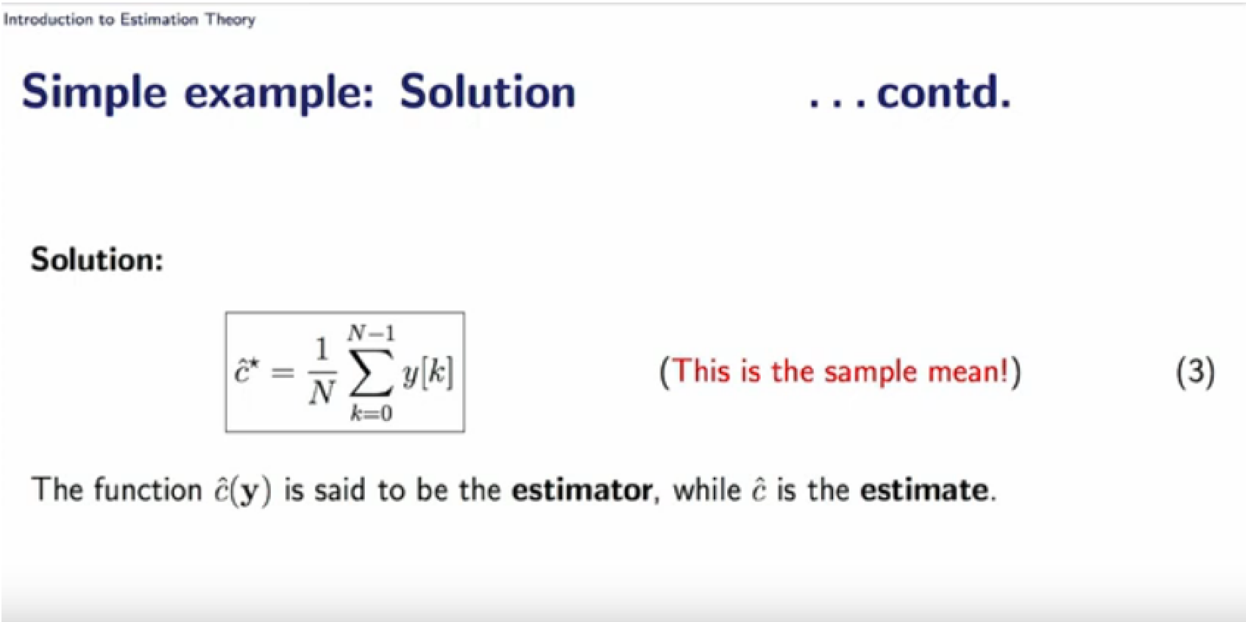
It turns out that, the solution is sample median. So you see the beauty behind this systematic approach is that I know clearly, where I have the freedom in estimation. In what aspects do I have the freedom? What are the different aspects in which I have, the user has a freedom? Choosing the model and choosing the objective function. But as I keep changing them, the estimates are going to change. That is optimal estimate is going to change. So you see, the moment I change my criteria, the solution is changed to sample median. Isn't sample median quite different from sample mean? In what respect it is different? Can you name one difference? Whatever you see? Whatever you see as a difference, tell me. Don't tell me the spelling Apart from the spelling. What can you say? Do you know what sample median is? You sort and pick the middle value. Do you see first of all there is a big difference between sample mean and sample median? What is the difference? What about you Rashmi? Can you name one difference? Mean is robust or median is robust? Why? First of all you have used the word robustness, which we have not yet talked about, but I understand what you are saying. The second you have said is minimum mean square, that is not correct, it is minimum some square, prediction error. When we talk of minimum mean square, typically this is what we mean. It is minimum some square, prediction error. Okay? But the fact is mean is not robust, median is robust. Yeah, that is it, so that is true sample median is more robust. What do you mean by robustness? Sorry? Out layer, if your data contains and out layer that means some extreme value, that extreme value is going to hijack the mean because it is very sensitive, it has simple averaging. In robust statistics there is a term called break down point. Our cricketers also have right? One short ball, out of the pitch they go back to the pavilion. That is where the cricketer breaks down. Say "oh my God again a short ball on my head, I am out". So you can now classify cricketers into robust and non-robust. Seriously, there are cricketers who know how to handle short balls; sample median is one of that. You add any extreme value, the value is not, the sample median is going to remain unperturbed for a long time. That means, for the quite a range of extreme values. Why is that because you are sorting and picking the middle value? So the nature of the operation is like that. Okay so, robustness, yes there is a big difference between these two, but that is important to observe, by changing the objective function, I am able to impart a different property to this estimate. You, see now I know, if I want to give some property, to the estimator, where do I have to meddle? What do I have to meddle with? The objective function. So the objective function plays a tremendous role in the estimation. Secondly, Sample mean is the linear function of the observations, it is a linear estimator. We wrote here right?

Refer Slide Time :(09:03)



Theta hat is a function, so what is sample mean? Sample mean is C hat, one over N, sigma YK right? Or you can say one over N, this one vector transpose, times y vector. Slowly you should get used to the vector form. It is a linear function. What about median? It is a non-linear, why? Because of sorting. Sorting is a nonlinear operator. So I have sample mean which is a linear estimator, sample median is a nonlinear operator. That is another property of the estimator. You may ask "what is a big deal, if it is a linear or non-linear? the computer is going to calculate it, Why are you wasting my time on it?", Well it makes a big difference, may be not so much in this example, but when you are going to implement this estimator online, what is preferred? Computationally which is friendlier? Linear. So through this simple example, hopefully you understand. How typical estimation problem formulated, what is the procedure now? You have data, so it will be clear on what information you have, in this example, they information available is the data. In some other example, information available may be data plus some aprioryknowledge of what? Of the modal. You talked about, black box modals, Gary box models in sysaid; typically we bill black box modals; that means the user postulate model. But many times I may have aprioryknowledge of the modal, in which case, I have my information data plus this aprioryknowledge. Then, I move on to the modal, include all the constraints followed by what? Followed by choosing my objective function. Then the solve the problem, out comes your formula estimator, here we able to derive, analytical expression. But many times your objective function, you may choose it to be Li squares, but it may be non-linear Li squares, here we your solving the linear Li square, what is different between linear Li squares or non-linear Li squares, on what basics are classifying, whether the predictor Y hat is the linear function unknowns or not, is it a linear function of unknowns example, yes or no, it is so its linear Li squares. Right? In many cases, may be non-linear one. So may be solving the non-linear Li squares problem, for which is this no clots from solution. So you have to resort to a non-linear solver, numerical solver, numerical optimization. So there are many fascist of estimation, now, now comes the biggest thing. Okay? We have arrived, we have learn to estimate grate, but we have done this for a single realization correct. The lies I can use this C star or

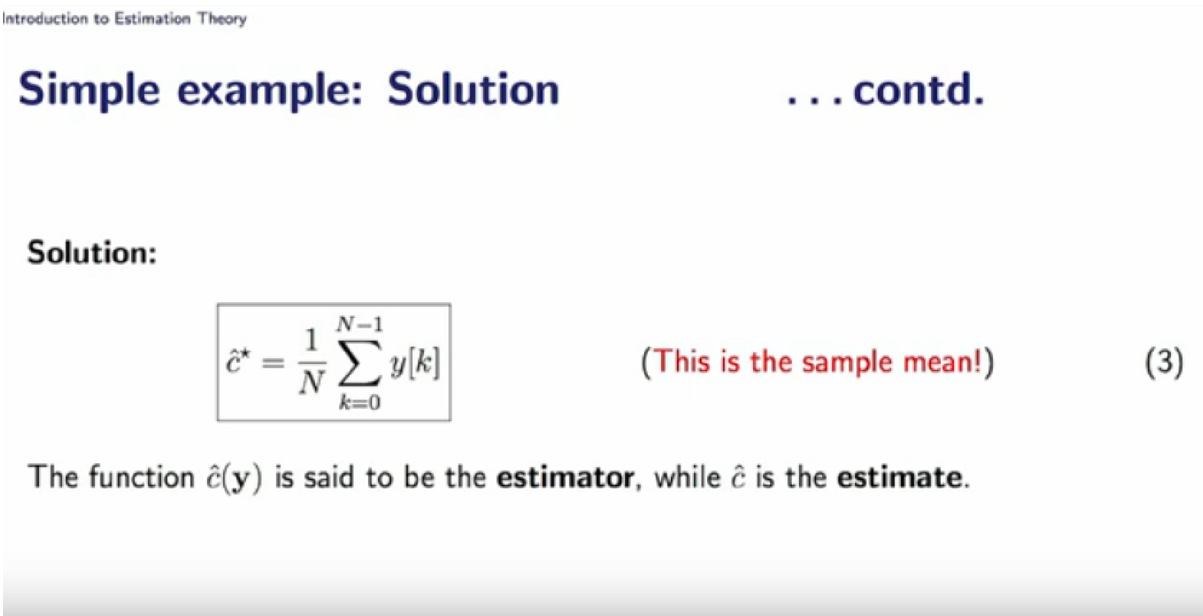
Refer slide time: (13:28)



I can use, this sample mean, of course spoken about is estimator sigma square E, but we will talk about the latter. Where I know very well, if I feed one data record, I will get one C hat, another experiment will generate another C hat and so on. This optimization dyson guarantee, the buyers for the example is going to be zero, I don't know, I had to separately examine that, that means, how this estimator is going to respond different realization said do not know, all I know is given a single finite length realization, How to optimal estimate, you understand. But how it response to different realization, remains to be assessed. Okay? And in that respect, we have to recall the diagram schematic that we do yesterday, data go through, goes through this estimator and producers, which in term producer's theta had. Right? So I want to now, ask the question, I would understood, how to bill in the estimator through this example, of course, we will apply this later on to our modals. But, now I had to ask a generic question, given an estimator, how do you know, if it's unbiased?

How do I know how much precision that means what is a variance in, in fact I wander and estimator we set earlier, which kens minimum variance, we call such as estimator, efficient estimators, why because one think you have to understand is that data come is with some uncertainty here. Data as some uncertainty and the in this example. We can set this uncertainty this measure sigma square like correct? Data some variability, this variability going too propagated, to variability in theta hat, through what through is a estimator. So, I have here sigma square, theta hat you recall in Fourier transform set you can view, Fourier transform in two different waves, you can think of it as signally composition or enrage of power leek composition here also you think of estimator as data in theta hat out, another way of the looking at is sigma square in sigma square theta hat out current, what do you I want I want sigma square, theta hat to be as a low as possible I am Right? It is a correct or not? I want to variable it in theta hat, to be as low as possible what this is a variability, again tell yourself, over oak rock is experiments, repeated experiments. Now different estimator for the same process can give my different, sigma square theta hat. Right? Then I will pick the one the fill me the lowest, sigma square theta hat and the one that gives me lowest, sigma square theta hat, it's called the,' Most Efficient estimator', it's very efficient it able to really chenout the uncertainty in why hat and reduce, the propagation of uncertainty as mush possible, if follow. So, let me asking simple question and one more point we a jump suppose, I will to ask you if this estimator here if it's unbaiasedness how you calculate? How do you verify analytically suppose ask you, this is sample mean is unbaiasedness. So, your see hat equals one over n sigma YK, easy it unbaiasedness. How do you calculate all you after do is exportation of theta hat, what is theta hat here, for us one over and sigma, sigma YK.

Refer Slide Time :(18: 11)



So, Exportation of this, minus one minus C not. Right? Where C not is a truth, are you can say mu, you C not is mu, so if assume that, Exportation of YK, if YK being generated by C not or mu not, whatever you want to call? Let's call it as, 'Mu Not' what will be your answer? What do I want? This difference to be, zero, is it zero? Yes or no? Very simple problem no, Exportation operators or liner operator, so how do you evaluate? Exportation of one over and sigma YK, you can take Exportation inside the summation: that is beyond a summation, what is exportation of YK? Mu non, n times mu non, divided by n that is mu non, so mu non is mu not is zero, so sample means is unbiasedness estimator, very good, but I only it tells me about the unbiasedness, it doesn't tell me about the precision, how it varies, house C hat varies across experiment I do not know, we will derive the expression tomorrow, but, I need to know that, so what we have just done is we have analyse the properties of the estimate. But, yesterday I said something else, what is that? What affects the quant, quality of the theta hat; one is the estimator, which is the processer, other is quality of the data, what you’re feeding in Horsham matters? If I feed in garbage and expert estimate to miracle’s it won’t, you can’t expert garbage in diamond sound, if you feed in garbage estimator produce another garbage. We want that some diamond plus garbage in and diamond sound. Right? So, first we will ask, how do we assess the quality of this data and that is done through, fesses information. So, tomorrow will get is the will focus on fesses information and then move on to buyers and various and so on. Okay?