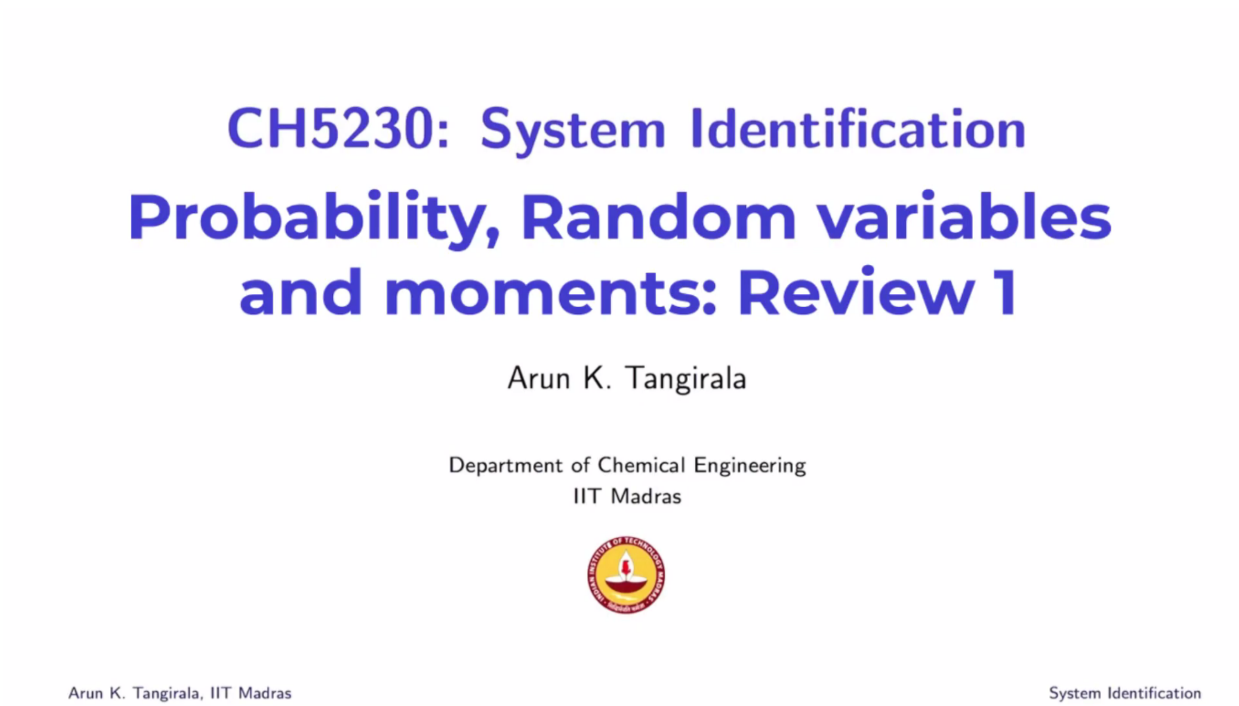
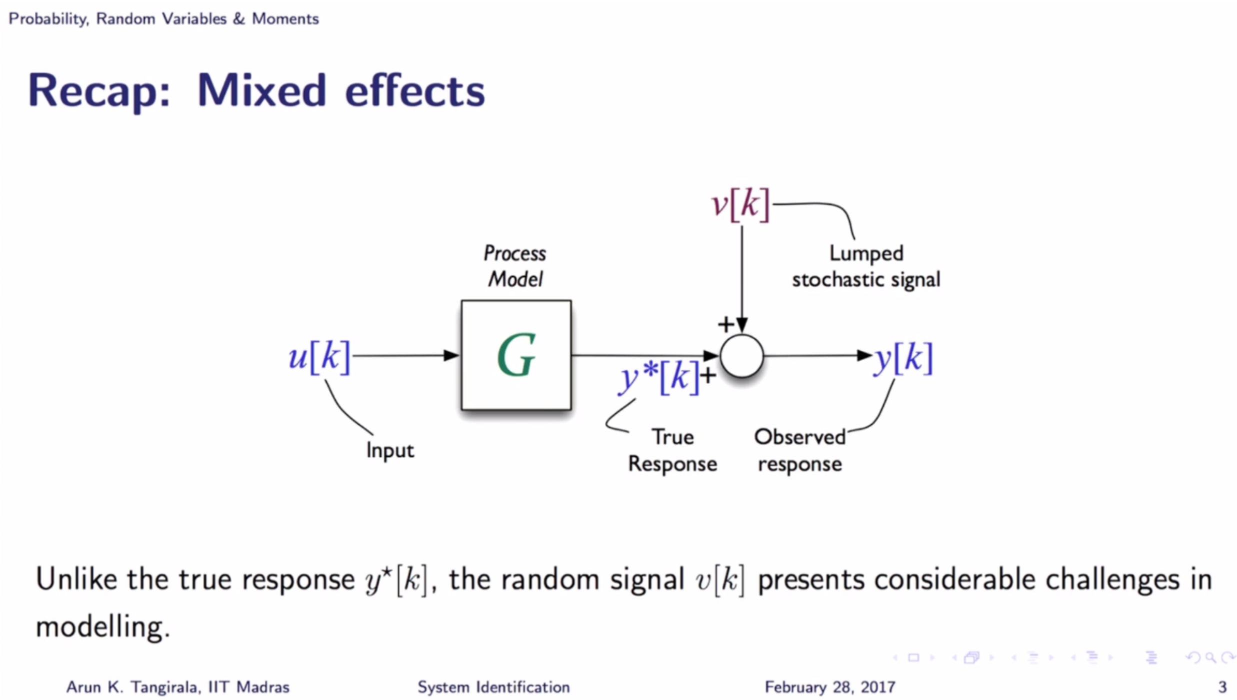


In the last class we kind of brought a closer to the discussion on the theory of deterministic linear time and variance systems. And I just gave you a very quick sort of opening remarks on random processes but I thought let me do it all fresh when it comes to reviewing this theory of random processes.

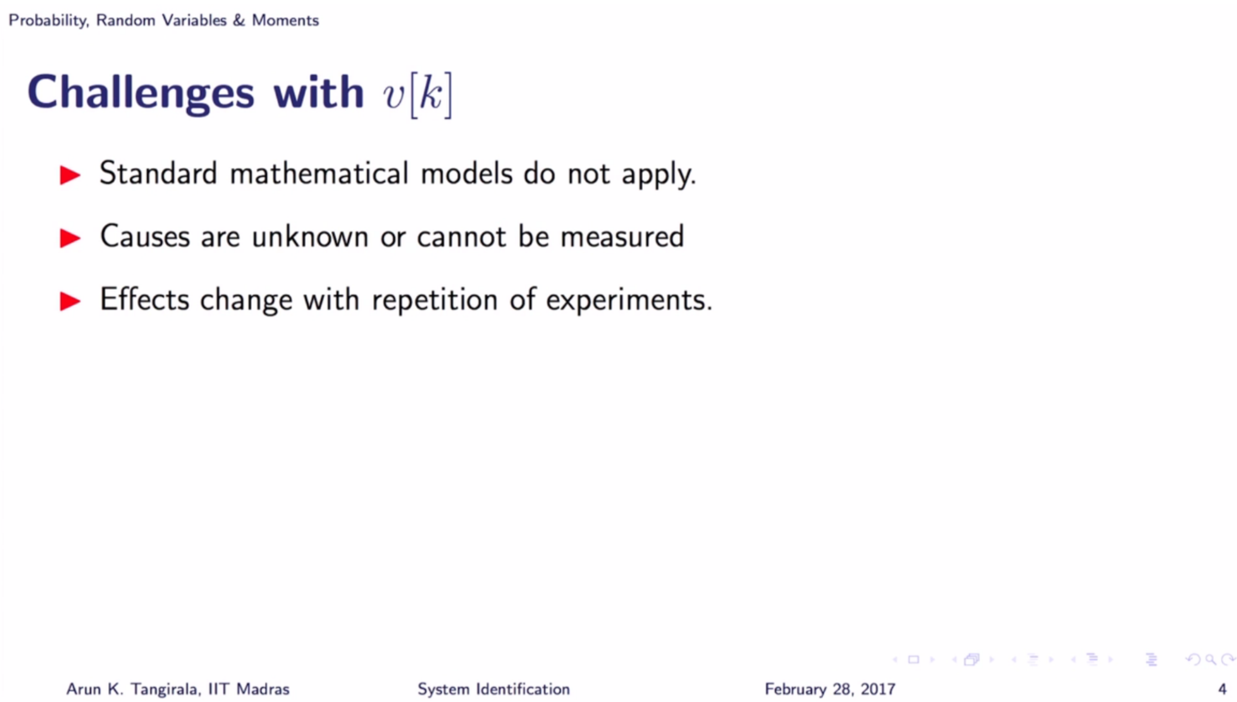


So from today's lecture maybe for about next three or four lectures maybe we will be reviewing the theory of random processes. I know that I am reviewing almost an entire semester's course on time series analysis in about four lectures but as I have said the video lectures for time series are available online and for those of you who are not familiar or who need to really go into detail in some of these concepts you will need to but those of you who need to can go and listen to the detailed lectures. I am going to give you a review of the salient points that are necessary for identification. But the most important part is that you should be comfortable with applying those concepts, understanding is one part of the story, applying is another part of the story. So let's get stated. Now if you recall as far as system identification is concerned we are looking at a composite signal or a mixed set of effects; deterministic plus stochastic and we assume of course that these stochastic term adds on to the deterministic one as I have repeatedly mentioned and this stochastic signal VK that you see on the screen is a lumped effect. Until now we have focused on the G part, that is the what is known as the process model or the plant model, we took that apart and we said okay let's look at this sub-system and we have learned how to express G in different ways, mathematically different ways starting from convolution to states based on and also frequency domain representations as well.

So unlike the true response Y star which was fairly easy to model. Why I say fairly easy to model is all you had to assume G is LTI and you could write the convolution equation and you can take off from there. VK which is this lumped stochastic effect and you have to keep asking yourself what is that we have lumped together. What is that we have lumped together here in VK?

Excellent. Good. Good. So sensory noise and effects of measured disturbance and in practice VK will also contain modeling errors. So theoretically VK contains the effects of sensor noise and effects of disturbances. But in practice V will also contain any modeling errors that we make in G.

Now this VK presents considerable challenge to your Y star because this is a stochastic signal. It's not your deterministic signal and what is so difference between a stochastic signal and a deterministic signal is something that you have to slowly get feel of and I will point out those differences at appropriate time in this review.



Now the first challenge that we have VK is that there is no standard mathematical you can straight away apply. Like I can't use maybe some non-linear version of some equation or any of the equations that I have learned and so on. There is no straightforward and as a standard here what I mean is straightforward I don't have any straightforward mathematical models.

And the second challenge that I have is that the causes are unknown. I don't know caused VK. I don't know why sensor gives me error. I can't go and question the manufacturer. It's like the story that we used to read. Why did you drop the milk you ask a little girl because ant bite me and you ask the ant why did you bite well because I was feeling afraid, or something was chasing me and then you go on. You will never be able to trace the cause. So those short stories that we hear in our childhood that's the case here.

We don't know why sensor is giving me noise. We don't know why there are disturbance. There are going to be disturbances. I can't really say no to those disturbance.

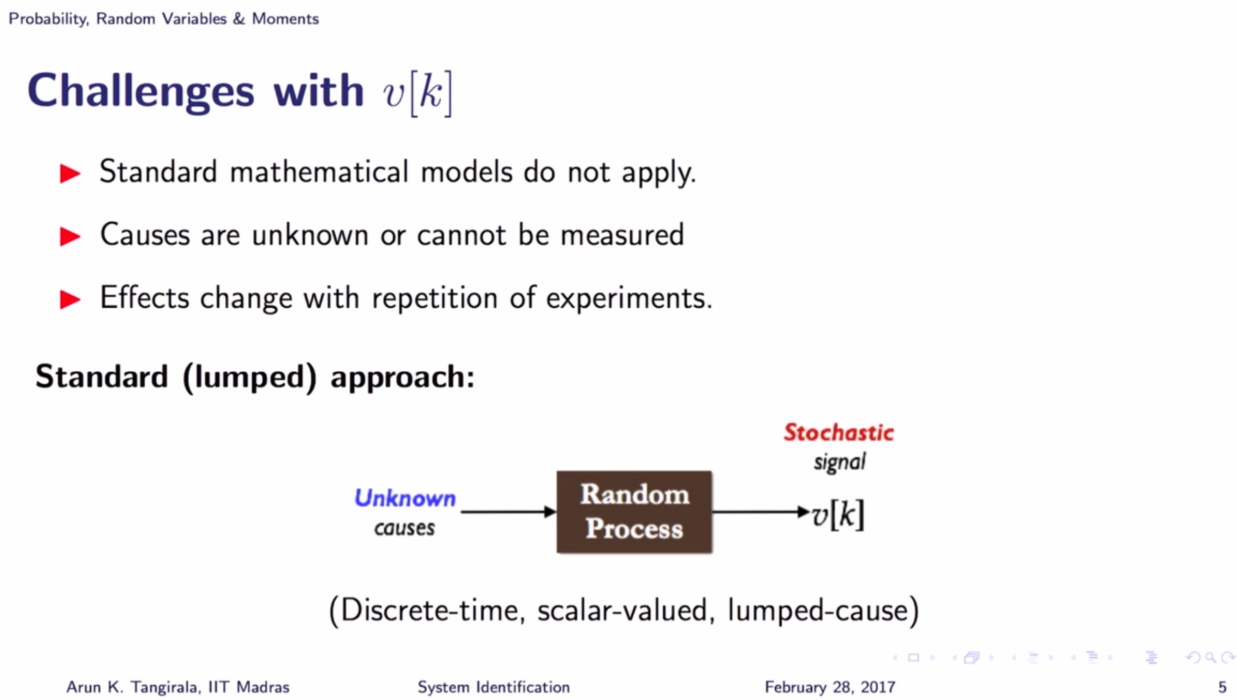
So Y star we do know that the source is U that is not U but the input U. and that there is a mathematical model that will allow me to describe Y star whereas VK I don't have a straightforward mathematical model. I do not know the causes even if I know I cannot measure. That is another difficulty to top this. And the third challenge is a feature of stochastic process or random process which is that every time I perform this experiment Y star remains the same but V keeps changing. Okay. That means every time I observe this process I have a different so called realization of V and that doesn't happen with Y star. Why do you think that's a challenge? Where do you think that's going to pose challenge in identification? So you say so what I mean do I care really?

Right. So from the data that I have suppose I know anyway to input U suppose I have a model for G I can discount for the effects of U and say yeah this represents V and I go ahead and build the model. But that model will be specifically for that realization. If I choose to really give a lot of importance to that. The third point tells me you have to model V if possible but keep in mind that you cannot give undue importance to the values of E. You understand if VK is a stochastic signal it maybe predictable I mean we know that it's not accurately predictable. That's a feature of a stochastic signal but it doesn't mean there is no predictability at all. So I want to see if I can build a mathematical model which will predict whatever is possible to predict but at the same time keep in mind that the values of V that I have are only specific to this realization. Had I performed another experiment, had I used maybe another sensor I would have obtained different values of it. Then in which case I have to be careful here. I cannot fully trust those values there. I cannot place complete faith in that which means that there were many many possible realizations of which I have observed one, and now I have to take that into account also during modeling.

That's a prime difference between dealing with a deterministic process and a stochastic process. When I have the response of a deterministic process whatever I see is a truth. There is no other possibility. Whatever it is it is. if I have a deterministic system I excite with the sine wave whatever I see as a response is a response. There is no or maybe that, maybe this, and so on. So I can just go ahead and build the model whereas with the stochastic process first of all I don't know the cause, I only have the response with me. And I can't even fully trust this response. I can't say these are exactly the values that I should be using in my model. I have to use them but not use them both.

The other way of looking at it is from a single realization of this stochastic signal I have to get a feel of what all the other possibilities are. That way I mean that is another way of looking at it which is of course challenging. But it's not such a griming situation after all. It is challenging but there are ways to overcome this but not fully though which means straight away we know that the quality of the model deterministic model that I can fit is going to be much better than the quality of the noise model that I am going to fit. Obviously because I have only one realization of the many many possible signals VK. So I have to like touch and go whereas with deterministic model I can spend time. I can make friends, it will stay with me for life long and so on.

So the quality of models typically noise models are not as great in any time series modeling as with your deterministic models. Now this standard approach to dealing with these challenges in VK is to use a lumped approach. As you see on the schematic there, in the previous schematic we only said that VK represents lumped effects. But now we are going step ahead and that is the standard approach in time series analysis to say that this VK which represents lumped effect now I will imagine.

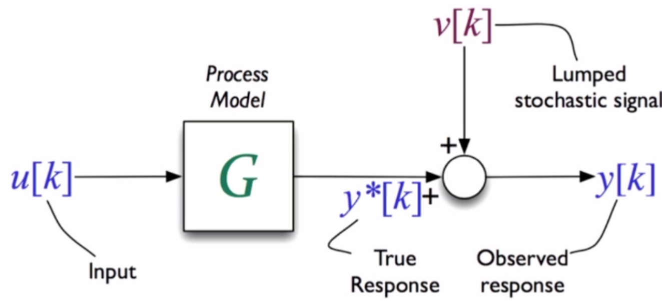


Now I am bringing in my imagination. There is nothing wrong with bringing imagination. After all the entire modeling world is imagination only. As long as you do two things, show that this imagination works, for fairly large class of conditions and two, that this imagination that you have you state under the condition, the conditions under which it will work.

So first you have to say when your imagination works and also show as a consequence that whatever conditions under which this imagination that we have is not too restrictive. It should not apply only to your process. It should apply to fairly large class of processes. So what is this imagination? This imagination is that VK which is your lumped stochastic signal so right now in this schematic I have thrown away Y star. I am just looking at VK is being driven by some unknown random signal. Now that is an extremely important abstraction in time series analysis. It's an extremely important abstraction. You may wonder where out of the sky is abstraction has come in, why should this diagram give you the impression that VK is being driven by something else? But if you come to think of it you say that well I don't know what the causes are. I don't know what is – something is driving VK for sure.

Probability, Random Variables & Moments

Recap: Mixed effects



Unlike the true response $y^\star[k]$, the random signal $v[k]$ presents considerable challenges in modelling.

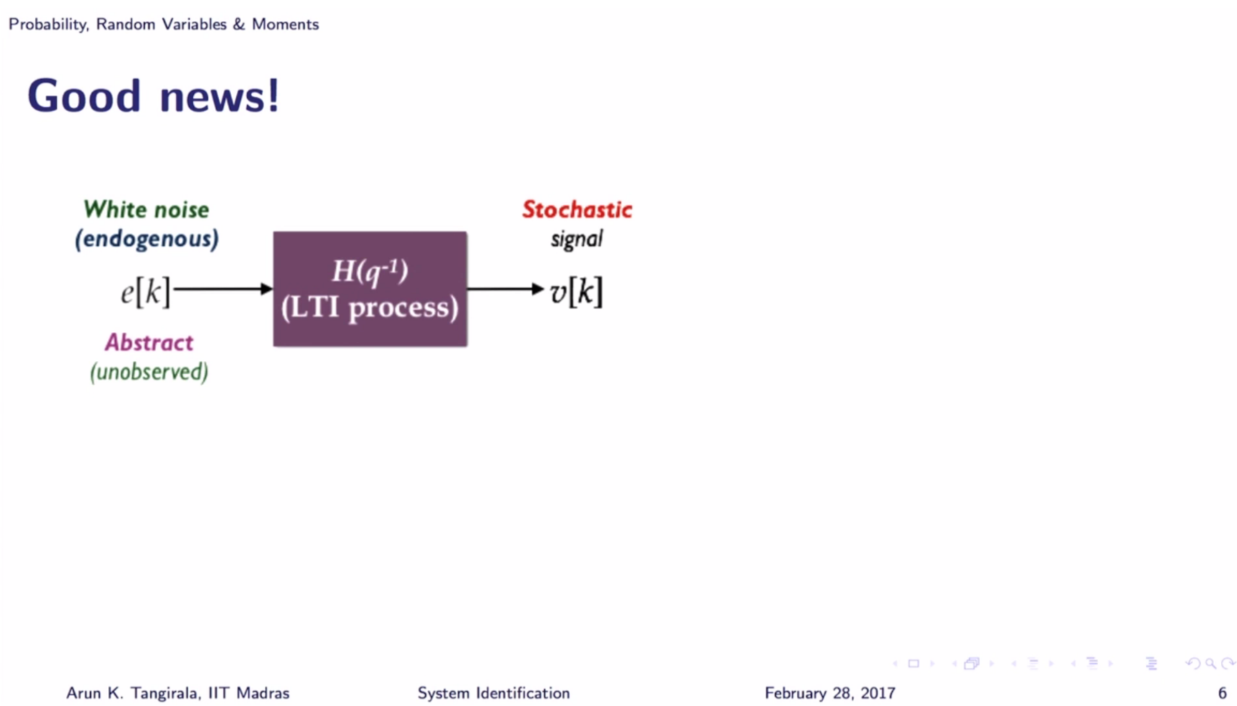
Arun K. Tangirala, IIT Madras System Identification February 28, 2017 3

I am not able to measure it. So that I am going to call as some unknown cause. And very soon we will realize that's white noise that is what we call as a white noise process. And I will show you at a later time why this imagination is justified. When this is justified we will learn very quickly but why does this imagination – why this kind of abstraction comes out. So the process that connects the unknown cause to the observed effect is what we call as a random process. That is one way of defining random process. There is a formal way of defining a random process which we will also review shortly. But this it the basic idea here that we are going to assume VK to be driven, to be a result of some random process being driven by an unknown cause. And a very important caution that I want to give you here this schematic may give you the impression that this unknown cause that we have or causes that are lumped is external to VK. Unfortunately it isn't. It is just endogenous. We are in fact when you look at the equations and later on when I talk about justification, for the abstraction you will realize that this unknown cause is actually is not external to VK. It is endogenous to be. You call that's endogenous. It is within VK. It is self driven. These are called self driven processes. There is no external cause per sey. There maybe in reality but I don't know what it is. I don't know. So as far as a modeling part is concerned this unknown cause although schematically we represented as some external signal it isn't external to VK. Unlike Y star and U. If you take Y star U is external. It's an external signal. In fact in time series literature, terminology, U is called exogenous signal whereas this unknown cause which we will soon represent as EK white noise is endogenous. And time series modeling preceded system identification so these models such as auto regressive models, moving average models existed much before system identification came in. and when somebody said now I have a problem the one that I have in system identification they said it is actually an auto regressive model with exogenous effects.

So for time series community the effects of inputs are considered exogenous. And that's how the name ARX model was born. Auto regressive model with exogenous input. RMA model with exogenous R max and so on but box although contained an X that X is not exogenous. Box is the name of statistician. Incidentally his name also ended with X.

But that X in Box maybe you can say excellent and extraordinary. He was a great statistician. So the Box – Jenkins model that we will learn later on also can be thought of as an time series model with exogenous effects. Very good.

So what we are looking at is a discrete time random process, scalar value that means we are looking at univariate random processes and lumped cause. We are lumping all of them and this cause is not external. We are assuming VK to be self-driven. We have just pulled out the wire of VK and so what is the special about this unknown cause. This – there is something special about this. And that special thing gives it the name white noise. What is that special thing? It is uncorrelated in time. See if you come to think of it how am I going to build the model for VK for predicting VK, there should be some dependence. That is if I look at the evolution of VK the past of VK should driving the present and the future. Only if such a situation exist I can build a model for VK. If there is no dependence at all in the history of VK or if you look at the direct evolution of VK there is no hope as far as predictability is concerned. Mean is the best prediction. Average is the best prediction. When we say I can build the time series model what I implicitly mean is that there is a prediction better than the average that is the implicit meaning. You should remember that. And such signals which lend themselves to building time series models are called correlated signals. We will talk about correlation very soon.

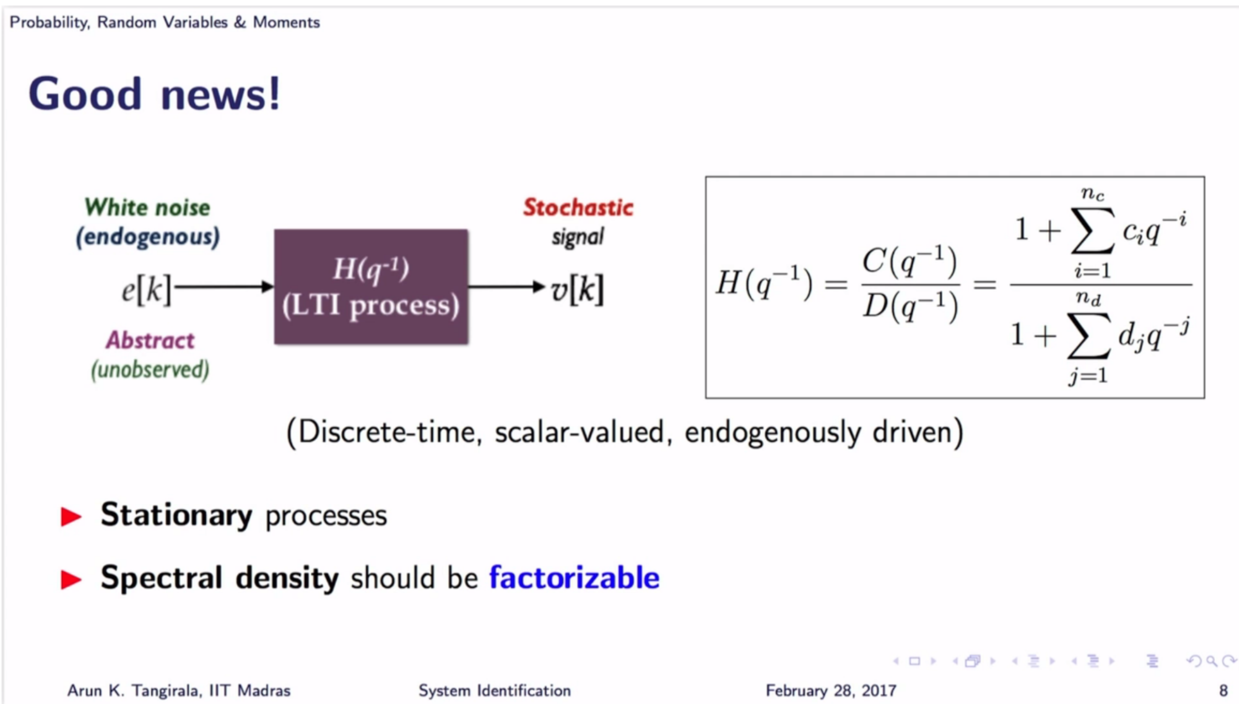


Now this unknown cause that we are talking about that is going to drive VK that doesn't have any correlation. It's an unpredictable signal. It's like your shock wave. You can think of it is an earthquake, whatever it is. It is the unpredictable process and unpredictable in a linear sense. You should remember that. What we mean by linear sense is that you can have predictability correlation and there is something uncorrelated and then you can have something called independence. All we are saying is that this white noise is uncorrelated. There is a stronger statement which is usually made called independence. And that will become clear very soon, what is the difference between uncorrelated and independence. We should not use them interchangeably in general. And as I have marked here this white noise are endogenous part of VK and let me write that here. So I have VK. I am given that it's a stochastic signal. We will break this up into two portions. A predictable portion given all the information up-to the past. This is the standard notion that we use. We had – the way we read it is V hat of K given information up-to K minus one not just K minus one and by definition in fact by old definition a stochastic signal always has an unpredictable component. You can never predict it accurately. We will denote that unpredictable component by ek. We don't know what is V hat made up of. We do not know whether V hat actually exist in V. that is whether there is a predictable portion or not but what we know for sure given that VK is stochastic this is always present. Always. V hat maybe zero or not, we don't know. That depends on the signal. If V hat is not zero then we say it's a correlated signal. That means there is some hope for prediction, some scope for prediction. If there is nothing in the past that will help me improve the prediction beyond the average the we might as well say that V hat is zero and VK is ek itself. So now you will see this ek that I have written on the board and the ek that you see on the screen they are the same. The first thing that you should see notice and I have deliberately used the same notion there. And now you can straightaway see that ek is an endogenous part. It's an indispensable portion of VK. So it is self-drive. The unpredictable part of VK is driving itself. We call as a self-excited processes like people that you see talking on the road into themselves these days you know that they are actually not talking into themselves. They have some cell phone and so on but they are just self-excited process like khud se pyar jagaungi something like that.

So that is the thing here. You have to understand unlike Y star that is what it is, and it's abstract. The only thing I know about ek is that it is unpredictable. What does it mean by that that if you assume ek to be zero mean. If you assume and this is generally assumed. So of course there are other conditions stationary and so on, we will come to that. I will straight away write in terms of expectation operator which we will soon see very soon. Typically we assume ek to be zero mean. What this mean – what we mean by uncorrelated is the prediction of e given all the information up-to K minus one is still zero. In fact strictly speaking it should be mu e the mean of white noise itself.

That means you give billions observation of white noise and you up-to K minus one then you ask well does this help you improve the prediction of ek beyond its mean, you say no. whether you give the past or not, the best prediction of e is average itself. So no hope at all practically. Whereas for V there is some hope. Roughly speaking this ek is like your shock-wave Quite often in newspaper you read market experienced shock-waves Where did the shock come from? You may be able to say but we say the experienced shock-waves Was this shock predictable? Not really. That really shook the market. So ek is that shock-wave that was not predictable. You couldn't see this coming at all. But it is endogenous.

Now the other important thing that you see in this diagram unlike the previous one we have just said unknown cause driving a random process generating VK but here we are being very specific, the good news is now that VK can be thought of as white noise passing through an LTI system. Why is that a good news? What is so good news about it? Until here we are kind of weeping VK is so difficult, there are so many challenges okay. I will think of it as being driven by some unknown cause and so on but now I am kind of happy, good news I am saying. No why would ek is still stochastic. Ek is still stochastic. So the repeatability is not the good news part of it. Any other reason that you should believe this is good news? Sorry. First of all I know how to model LTI systems. So I have all the paraphernalia already available with me. I don't have to separately study some other set of models for this. All the convolution, states, space, difference, equation everything comes here. The only difference is that earlier the inputs were deterministic. Now the input is fictitious and it is stochastic. The only thing I know about the input is it is uncorrelated. Why is it called white noise we will realize a bit later but for now take it as to be white noise.



So a part of the good news is that I already have the theory of linear time and variance systems with me. And I am glad to see that there are a class of the stochastic signals, by the way, not all stochastic signals can be represented this way. There are conditions under which this holds and those conditions are that first VK has to be stationary. And two that spectral density should be factorizable. I am not defined what is spectral density.

You are being bombarded with a lot of new terms today but remember this is review. So if you are not feeling too comfortable please go back and look at the videos.

So the first part is that it's an LTI process that is driving VK. And I know already that an LTI system can be defined like a transfer function object. In writing H of q inverse this way I have made some assumptions. What is that? Does an LTI process always have to be written this way? Ideally what kind of a model I should have written for H?

Student: [00:24:30]

Arun: There is no delay here either. If I tell you that there is a process that LTI what can of model comes into your mind? What is the first model that comes to your mind? Convolution. So ideally

I should have written here VK as sigma Hn ek minus n. Is this true? This is actually the definition of a linear random process. Of course not complete though, I will have to impose some conditions on H. This is actually the governing equation but I have written H of – I have written transfer function form. When I write it in this way, I am already assuming a parametric form. Only when I think of a parametric form I can write transfer functions and so on. So ideally speaking I should have written this convolution one. Same. In place of G you have H. In place of U you have E. And in place of I you have V that's it. But otherwise, things don't change. But it makes a huge difference now. What is the big difference between this equation and the other convolution equation that we have seen for G? So let's write below here symbolically or structurally they may look the same. Like this spot these six differences contest. You have to now figure out what are the differences. Of course in place of H you have G, alphabetically it's easy to remember. G comes first and then H comes later. And then in place of U in fact I should write here Y star specifically. Earlier we have been using Y but now we will be very specific because we know now that Y is a mixture of Y star plus and V. What is the main difference between these two equations here? U is known it's deterministic. This is unknown. It's fictitious. The only thing I know about this is that it is uncorrelated. That's all I know. We call as a statistical property. I only known the statistical property of it. I don't know anything else.

So think of now, building a model for V versus building a model for Y star, sorry for H or for Y star. When it comes to Y star the input is always given. So estimation of G is easy. Isn't it? But when it comes to building a model for E, am I given E? I am not. So how am I going to identify H? My goal is to identify H. we can still do it. The good news is we can still do it without knowing E just by knowing the statistical properties I can do it. Like there is a small challenge between and there is a small difference between identifying H and G and that difference again stems from the fact that I know the input U and I do not know the input E. I said that it is possible to identify H without knowing E. that is correct. Nevertheless, if you just look at these two models there is some non-uniqueness about this model. Let us just for the sake of discussion restrict ourselves to [00:28:27] models. Even then you have a non-uniqueness issue. What is a non-uniqueness issue in the model for V? Is that a non-uniqueness issue? What I mean by non-uniqueness is if I have a model let's say set of coefficient set that I have identified I can have another set of coefficient that will model the same process which means there is an identifiability issue. I have to fix that. But are you first convinced that there is an identifiability issue.

What does it mean by modeling –building a model for Y star I am given Y star and I am given U the goal is to estimate G either using an FIR approach or a parametrization approach. That's okay. That we will not worry about. Here what is the statement? Given V alone suppose you are given V in fact in [00:29:37] you are not even given V. You are given Y. But let's say I am given V. The goal is to identify H. Is that the only goal? There is an additional objective which you have to understand. So here there are two objectives here. Estimate H the sequence of coefficient and there is a second objective which is estimate the so called variance of e. It says stochastic signal. It has a mean, it has a variance and so on which I will – concepts I will review shortly. But we know from here say at least I know that a stochastic signal has something called a mean, something called variance and so on. I do not know what its variance is? All I know about ek is it's unpredictable. That's all I am given. And it turns out that is sufficient as far as estimation is concerned but identifiability is still a challenge. What I mean by that is if I have let us say a set of coefficient, I have a solution let's say I have a h hat of n and sigma square e. So I have this as one solution. I have ran an algorithm a times series modeling algorithm I have got my estimates of H and sigma square e. there is ye another solution that will produce the same V. You agree.

Do you agree that that there can be another solution that produces the same V? Do you see that? You don't. Do you see or you don't?

How many of you see that there can be another solution that satisfies this equation?

Student: [00:31:40] there can be several values of ek.

Arun: That's okay. I am not worried the values of ek. My goal is not to estimate values of ek. Do you see that there can be another solution to the same problem theoretically itself. Forget about estimates. Theoretically itself I can rewrite this model as so I can say where h tilde is not necessarily equal to H and E tilde is not necessarily equal to U. Both e and e tilde are white noises. Both are unpredictable. Do you think it's possible to write another model like this? Do you see the existence of another model is what I am asking. If that is possible then that it means it's not identifiable. That means there is no unique answer. What is the simplest other possibility?

Is it possible to write like this for Y star? Is it possible to write like this for Y star? There exists two sets of impulse response coefficient that are different from each other. I am not changing the input. Can I change the input? Do you see the difference? Can I change the input for the deterministic system? I didn't write U tilde Why is that? U is known. It's fixed. I can't change the input. The only thing that I can change is the model but for a given LTI system and given input there exists only one. G has to be equal to G tilde whereas here the input values are unknown. They are fictitious. So I can always think of another input which has a same statistical not necessarily same statistical property but it's also uncorrelated. It's also another white noise sequence but they may differ by a factor alpha. So which means I can have h tilde as let us say alpha times h and when I change that I can have one by alpha, right?

So that means e tilde will be one by alpha e for any non-zero value of alpha. Where do you think this non-uniqueness is coming from? What is the cause? What is the reason that I am able to write two different models for the same V, both are LTI. What is the prime reason? Input is arbitrary, abstract. Only something I know about the input but that is not about on the values. I only know that it is unpredictable. But that alone is not it is sufficient to build the model but it is not sufficient to build the unique model. Get that. I said earlier I can estimate h exploiting the white noise nature of e that's not an issue. But I cannot guarantee the existence of a unique model. For that I have to fix and one of the fixes that is unusually done is that in this noise modeling to fix this non-uniqueness the first impulse coefficient is set to one.

That's a fix. That is not the fix. It is a fix. That means there are other ways of fixing it. Fixing this non-uniqueness. So when you expand this, with that fix what do you see? What do you see as a model for – okay. Did someone get me the answer? Okay. So here. When I expand this with this fix what is the first term? Ek, good and then the rest right. Ek plus sigma hn ek minus n with n running from one to infinity. Do you see that? The advantage of fixing h zero to one gives me easy interpretation I can straight away see ek as an integral part of VK. I could have fix it to two, nothing prevents me from doing that. But it is just universally fixed to one. There are reasons also of doing that. Let's not get into those reasons. But you don't have to do any of that in the deterministic modeling. And now look at the transfer function that you see on the screen. Do you see something strange about the transfer function compared to what you have seen for Gs? What is it? You have to tell me what is the difference that you see for the transfer function h as against what you have seen for G.

Student: [00:37:43]

Arun: Sorry? First of all is there a delay? So there is no delay. Why is there no delay you have to ask yourself. Why am I not incorporating a delay? Unfortunately, as we will learn also later on hopefully at some point in time or you can refer to the time series lectures, even if there is a delay between the shock-wave and VK I cannot identify it because ek is unknown. Input is unknown and then on top of it delay is unknown then what delay are we talking about? I can think of delay only when the input is fixed. If the input is not fixed I can always adjust the input and say this is the input. So that the delay is not there. So that means in time series modeling delays that is univariate time series modeling, please don't think in all time series modeling that's the case, in univariate time series modeling I will not be able to find out whether there is a delay between the shock-wave and VK. And it doesn't make sense also because the shock-wave is an integral part of VK. What do you mean by delay? It's there with VK all the time. So that's the first difference that we notice between the G and H. What is the second one you mentioning something?

Student: [00:39:05]

Arun: Correct. And the denominator anyway begins with one because the coefficient on Y star is always one. Here the first coefficient in the numerator is also one. We call them as monic polynomials. The numerator and denominators are called monic polynomials which begin with the coefficient unity. Why is that? Why is that the case? Because we have fixed the first impulse response question to one. Do you recall how we derive impulse response coefficient given transfer function? Long division? So only when we have anyway denominator polynomial will always leading coefficient one. If I want the first impulse response coefficient to be one, the first coefficient the numerator also has to be one. Therefore, by default, all these noise models that you are going to build at least for linear random processes will always have a one in the numerator and in the denominator. Denominator anyway it's there. It goes without saying. So that is something that you should remember as a very distinguished feature of h as compared to G.

So let's summarize now. VK was this – is this random component of Y that is a measurement which cannot be explained by the input. Point number one. Two, it cannot be modeled as we model U because it doesn't fall within the realm of deterministic processes. Alright. And three, with all – that there are several challenges to modeling VK. And four, that there is a good news, which is that, despite all of these challenges I can still think of VK as being driven by a white noise passing through a filter. As being there output of a filter being driven by white noise. Let me put it that way. But with some terms and conditions like say T and C apply.