Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 02

Lecture 07: Basic Time Series Processes

Hello all. So, welcome to this new lecture in this course on time series forecasting with applications in R. So, again, if you vaguely remember where we stopped in the last lecture. So, the main aspect in the last lecture was time series decomposition, right? So, we'll try to finish that off really quickly, and then we can talk about some basic time series models. So, when you talk about any time series decomposition, you have further two types of time series decomposition.

So, the first one is called additive decomposition. So, what exactly do you mean by additive decomposition? So, the idea is fairly easy: if you can actually represent any time series Yt as a sum of all the individual components, so let's say trend, seasonality, cyclicality, and then the last one is irregular components. So, Tt plus St plus Ct plus It, then one can actually say that the underlying time series Yt has an additive decomposition, okay? And then it represents an absolute amount.

$$Y_t = T_t + S_t + C_t + I_t$$

So, if you can represent the time series in an additive nature or in an additive manner, then one can say that it kind of represents an absolute amount. So, we can talk about a simple example. So, let's say if a manufacturer produces 10,000 more machine parts in November as compared to October. So, let's say in October, if the manufacturer produces X number of machine parts, and then obviously, in November, it will be X plus 10,000, okay?

So, this is an additive structure, right? So, we will say that the decomposition in the underlying time series of producing machine parts or machine part observations would be additive. Now, similarly, the second kind of decomposition in the same umbrella is called multiplicative decomposition. So, what do you mean by that? So, again, the idea is easy:

if the underlying time series Yt can be broken down into the product of all the individual components.

So, Tt into St into Ct into It, ok? Rather than a sum. So, the only difference is in additive you have a sum, whereas in multiplicative we have the multiplications. Now, obviously, since you are taking the multiplication of the product of all the components, it represents a relative amount as compared to an absolute amount. So, relatively, how much is the change? So, again, the same example in a slightly different way.

So, if a manufacturer produces 20 percent more machine parts in November as compared to October. So, let us say if in October the manufacturer produces 100 machine parts, then in November since you have to increase by 20 percent, it will be 120 and so on. So, here the nature is not absolute or the nature is not additive in nature; the nature is more multiplicative in nature, ok? So, this is one small idea of a slightly different way of decomposing a time series, which can also happen. So, either in an additive or in a multiplicative kind of context.

All right, so now the next thing we'll talk about is really basic kinds of some time series models. So, we'll describe three or four time series models which are more like building blocks to any kind of advanced models that we'll see in the next lectures. For example, the first one is called a white noise process. So, the white noise process is a very famous process and one which is quite used in a lot of application areas also. Now, initially, I would like to fix some notations.

So, I would like to keep the same notation throughout the entire course. So, here if you see, let e1, e2, e3, e4, etc., be a sequence of some random errors. So, what exactly are these e's? So, e's are random errors, ok? So, e1, e2, e3, etc.

So, the e sequence is nothing but a sequence of some random errors. Now, you have some properties of the random error. So, random errors have to be IID. So, again, remember what IID means. So, IID means independent and identical.

So, the distribution of all these errors has to be the same. So, let us say normal, exponential, it can be any distribution, but the distribution of all these e's has to be the same, and they also have to be independent among themselves. So, e1 and e2 are independent, e1 and e3 are independent, etc. So, IID stands for independent and identically distributed. And then we will try to specify some constant quantities.

So, let us say the mean of any of these ei, for that matter, is 0. And the variance of any of these ei is again fixed. So, σ^2e, okay. Now, one important point to note here is that neither the mean nor the variance depends on the time frame t, if you see, right? So, the mean is 0.

So, 0 is a constant. Similarly, σ^2e also does not depend on t, okay. So, this is a very important assumption underlying the white noise process, okay. Now, how exactly should we construct this time series, which is called white noise? So, the time series Yt, the underlying time series Yt, is nothing but exactly equal to the corresponding random error, okay.

Which is et, ok. So, I will say that this is the most simplistic kind of time series model which one can encounter. So, at any time point, let us say 1, 2, 3, 4, 5, etc. At any time point, the value of the underlying time series process Yt is exactly equal to the corresponding random error, ok. Now, imagine how this white noise process would look like, right?

So, since at any time point, let us say the x-axis is time here. So, since at any time point, you basically have a random error, right? So, again, this would exactly depend on the noise kind of structure which is completely random, right. So, for that matter, it is called white noise. So, the name also suggests that the time series process kind of resembles a noise time series, all right.

And again, why noise? Because, again, at any time point, you have a random fluctuation of the time series, ok. Now, more interestingly, why is it called white noise? So, the name white noise is interesting. So, imagine that if you must have sat through a physics course in your life. So, imagine that you have a white light.

And if you throw that white light on some surface, then it kind of distributes into all different kinds of seven colors. So, a broader spectrum where you have red, orange, green, blue, etc. So, imagine that the fluctuations of that white light kind of resemble exactly the noise which you see here, and hence this process is called the white noise process. So, the first basic time series model in our case is the white noise process. Now, moving on, the second model is slightly advanced and is called the moving average process.

So, I am quite sure that many of you might already have heard about this process, which is called moving average. So, in short, we will call it MA. So, the moving average process is called MA. Now, again, I will stick to the same notation. So, all the ei's.

So, e1, e2, e3, e4, dot dot dot, what are they? So, again, they are a sequence of random errors which are again IID. So, independent and identically distributed. With again some constant mean, so let us say 0 mean, and some constant variance, so let us say σ^2e, ok. Now, how exactly are you obtaining the underlying time series Yt or how exactly are you getting the underlying time series Yt is through this slightly advanced structure.

So, Yt is nothing but et plus one half et minus 1, ok. So, imagine that we go down the history. So, we can have t, then t minus 1, then t minus 2, t minus 3, etc. So, what do you mean by t minus 1 is nothing but the time stamp is just one time stamp in the history. So, the current time series observation which is Yt is nothing but the average of certain errors.

$$Y_t = e_t + 0.5e_{t-1}$$

Can you see that? So, the current timestamp or the current observation of the time series Yt is nothing but a simple average of the current error, which is et. And you're putting some weight. So, let's say half, one half, et minus one. But in a way, if you observe this entire structure, this entire structure is nothing but a certain average of errors.

So et plus one half et minus one. And imagine that if you roll this time window into the future, So, do you not think that this t would also keep rolling and then you will kind of get a moving average? So, I will give you an example. So, let us say if this is a structure, what is Y2? So, y2 would be nothing but e2 plus one half e1. Using this structure, right? Now, what is Y3? So, Y3 is nothing but e3 plus one half into e2, all right. Now, what is Y4? So, Y4 is nothing but e4 plus one half into e3, and so on, dot dot dot, ok.

$$Y_2 = e_2 + \tfrac{1}{2} e_1$$

$$Y_3 = e_3 + \tfrac{1}{2} e_2$$

$$Y_4 = e_4 + \tfrac{1}{2} e_3$$

Now, using all these individual processes, can you see that at every timestamp t, let us say 2, 3, 4, 5, 6, and so on, what we are essentially doing is we are kind of averaging the errors? So, the current error and the weighted average of one previous error, and we are also moving the time window. So, 2, 3, 4, 5, 6, 7, and hence this simplistic-looking process is called a moving average process. So, we are kind of moving the window, and we are kind of averaging the underlying errors also.

Now, the third kind of a basic time series model is also interesting, which is called a random walk. Now, again, if you must have sat through any stochastic process course, then this is a very simplistic-looking process called a random walk. So, what exactly do we mean by random walk? So, again, I will stick to the same notation. So, let us say e1, e2, e3, e4, dot dot dot, be a sequence of random errors which are again IID, so independent and identical.

With again some constant mean 0 and some constant variance which is σ^2e. Now, we have to pause here for a moment and just try to grasp the notations we have. So, let us say how exactly one can actually obtain the underlying time series Y2. So, initially at time point 1 where you are just starting, what is Y1? So, Y1 is nothing but exactly equal to the first random error, which is e1. Moving on, so moving down the timeline, how much is Y2? so Y2 is nothing but the sum of the first error and the second error, so e1 plus e2. All right, now how much is Y3? So, if you extend this further, Y3 is nothing but the sum of the first three errors, so e1 plus e2 plus e3.

Now, how much is Y4? So, Y4 is nothing but the sum of the first 4 errors. So, e1 plus e2 plus e3 plus e4, and so on. So, in general, if you want to write down the general time series process Yt. So, Yt is nothing but the sum of all the errors up to time point t. So, e1 plus e2 plus e3 up to et.

$$Y_1 = e_1$$

$$Y_2 = e_1 + e_2 \longrightarrow Y_2 = Y_1 + e_2$$

$$Y_3 = e_1 + e_2 + e_3 \longrightarrow Y_3 = Y_2 + e_3$$

$$\ldots$$

$$Y_t = e_1 + \ldots + e_t \longrightarrow Y_t = Y_{t-1} + e_t$$

So, in a way, I can actually write down this Yt in a summation format. So, Yt is what? So, Yt is nothing but the sum of the corresponding e_i, right, where i goes from 1 to t. Isn't it? So, one can actually reduce it to a summation notation. Summation of all the errors, all the random errors up to time point t. But, interestingly, what a few other people kind of use is that one can actually translate each and every row here in a slightly different manner starting from the second row onwards. So, again if you go back to the second row, the second row tells us that the corresponding Y2 is nothing but the sum of the first error and the second error. So, using this structure or using this equation, one can immediately get this equation. So, Y2 is nothing but Y1 plus e2, and why is that? Because Y1 is nothing but e1. From the first row.

Right. So, I can very well replace e1 here. In place of Y1. And get the same structure that we have here. Right.

So, Y2 is what? So, Y2 is nothing but Y1 plus e2. Similarly, if you go down the line. What is Y3? From the third row.

So, initially, we saw that Y3 is nothing but the cumulative sum of all the errors up to time point 3. But one can actually rewrite that slightly in a different manner. So, Y3 is nothing but Y2 plus e3 because how much is Y2? So, Y2 is nothing but e1 plus e2, isn't it? So, one can replace e1 plus e2 here and then plus e3.

So, at any given time point, what we essentially have is this. So, a general-looking notation is Yt is nothing but Yt minus 1 plus the corresponding random error at time point t, so, e t. Now, Yt is the current time series process, Yt minus 1 is one historical time series process, and e t is the current random error. So, this is a much neater-looking equation of a random walk. And essentially, what exactly do you mean by a random walk is that the current situation of the random walk is entirely dependent on one prior situation, which is Yt minus 1 plus some extra random error added to that, OK.

So, a random walk is nothing but the current state depending on one previous state only, plus some small random component which is added to that structure. Now, I'll give you one more interesting name for this entire random walk. So, firstly, we'll take up some examples. So, let's say if a person starts here at state one, right? So, this is my Y1, okay?

Now, as per the structure, what should Y2 be? So, if you use this structure, right? So, Y2 is nothing but Y1 plus some random error. Now, some random error means that imagine if you have a person who's sitting here, He or she kind of walks randomly in any direction because there has to be some random error.

So, imagine that if the person decides to walk here, and this would be my Y2. Now, again, obviously, this direction is completely random. So, let us say this is Y1, and this is my Y2. Now, imagine that the person is sitting here. Now, if you want to extend this, what would Y3 be?

So, Y3 would be nothing but Y2 plus some random error. So, let us say now I will take a random direction. So, let us say he or she comes down here, okay. So, this would be my Y3, and so on, isn't it? So, can you imagine this in your head? Imagine that you have a person sitting here in the first state, which is Y1, then in the next state, he or she can move in any random direction.

So, let us say he or she goes here, so Y2. Then in the next state, he or she can again move in any random direction. So, let us say he or she comes here. Then again in any random direction, so let us say it goes there. So, Y4, etc.

So Y5 can be here, right? Then Y6 can be here, okay? And so on. So, dot, dot, dot. And for this reason, you have a very interesting name for this walk, which is called a drunkard's walk.

And why drunkard's walk? So, imagine that you have a drunk person. A drunk person does not know where to move in the next step. So, he or she can choose any random direction and then go there. So, for that matter, a random walk is also called a drunkard's walk because a drunk person does not know where to go in the next timestamp, which is completely random according to how he or she walks.

So, for that matter, he or she can choose any random direction around themselves and then go there. So, this is the idea of a random walk. Now moving on, the next process is a linear trend process. So, probably in the last lecture, we saw what you mean by a trend. Or a general-looking trend, right?

So, now, we will try to describe this in a much more formal way. So, again, the same notation. So, let us say e1, e2, e3 be a sequence of random errors which are IID. So, independent and identical with constant mean 0 and constant variance σ^2e. Then how would you describe a linear trend? So, time series Yt is obtained as a plus bt plus et, where obviously a and b can be any constant.

$$Y_t = a + bt + e_t$$

Where a, b are constants

So, let us say 10 or 5, 2, etc. And again, this et is the same random error component. Now, clearly, can you imagine that since you have this time component ‘t’ here, right?

So, depending on if you move down the timeline, it would actually give you a trend kind of a structure. So, if you try to graph this, right.

So, since you have t here in the equation itself, it is entirely dependent on the time. So, let us say if the current observation is here, then the next observation has to be here depending on the sign, of course. So, if you have a positive sign, it will increase or be an upward motion, and if you have a negative sign, it will go down, right? But the understanding is that if you go down the timeline, then, since you have this t component here, it will go on producing some structure like this, ok, which is nothing but a trend or a linear trend basically, ok.

Now, I think we come to a very important idea that in any time series literature, you have two very basic models. So, the first one is called autoregressive, and the second one is called moving average, ok. So, we will briefly discuss what you mean by an autoregressive process. So, in short, any autoregressive process is called AR. And you see a letter here which is P in brackets.

So, we say that any autoregressive process has some order, and that order is P. So, P is a kind of general-looking order. So, we say that the complete model notation is AR(P). Now, a small definition to describe what exactly the AR process is. So, what we do here is we forecast the variable of interest using a linear combination of past values of the variable itself. So, we forecast the variable of interest, let's say Yt, using a linear combination of the past values of the variable Yt itself.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t$$

So, if the model uses the last p values, then the order becomes p, and hence it is called an AR(p) model. Make sense? Now, through the structure, you will be able to understand the idea of AR(p) in a better manner. So, what exactly is the structure? So, Yt exactly equals some C. So, C is nothing but the intercept that you see in a regression course or some constant.

It could be 10, 100, 5, etc. So, the current observation Yt is nothing but the constant C plus you have some coefficients. So, φ1, φ2, φ3 up to φp, and attached to all these coefficients are the historical data points or the past values of the variable Yt itself. So, Yt minus 1, Yt minus 2, Yt minus 3, and then up to Yt minus p, and then towards the end, you have a random error component, which is it. All right. So, hopefully, the structure is clear.

So, we are kind of regressing the current time point t of the process Yt with its own past value. So, Yt minus 1, Yt minus 2, Yt minus 3, etc. So, in a way, one can actually imagine that this is nothing but a regression equation. Where let's say this is the response, which is Yt, and all these variables, so Yt minus 1, Yt minus 2, Yt minus p, are nothing but the corresponding axis or the corresponding independent variables, okay. And hence you have a regressive component. The name itself, right? So, it is called an autoregressive. Now, the last question which remains is why is it called auto and not just a regressive process, okay?

So, the term is autoregressive. So, remember what is happening here. So, you have a single time series which is Yt. So, let us say Yt could be the stock price of Google, or Yt could be, let us say, temperature data, or Yt could be, let us say, sales data, etc. So, here the random variable, which is Yt, corresponds to one time series only or one application only.

So, the independent variables that we have here. So, these independent variables are not some entirely different random variables. They again constitute the same series which is Y, isn't it? So, Yt minus 1 is what? Yt minus 1 is nothing but yesterday's temperature.

Yt minus 2 is the day before yesterday's temperature, etc. But still, all these values kind of come from the same time series Yt. So, essentially what we are doing is we are kind of regressing Yt on its own past values. So, the term 'own' is important. So, own past values.

Right? And hence, it's called autoregressive. So, 'auto' means that since you are regressing the same structure Yt on its own past values, it's called autoregressive. All right? So, hopefully, this is clear.

The structure is such that the response variable is Yt. You are regressing Yt on its own past values. And of course, there are some coefficients, which are phi, that need to be estimated. So, we will talk about that later. But the structure of this autoregressive process is this. All right?

So probably, if you fix some orders, let us say 1 or 2, we get some basic autoregressive models. For example, the first one is AR1. So, AR1 is the simplest of the models. So, if you fix P to be 1, let me write it down. So, here P is 1, basically.

$$Y_t = c + \varphi_1 Y_{t-1} + e_t$$

AR (1) model is stationary if: $-1 < \varphi_1 < 1$

And then one can actually get this structure. So, Yt is the intercept plus φ1 multiplied by Yt minus 1 plus the random error et. Isn't it? So, again, if you go back for a second, this is the more general ARP model. So, if you fix P to be one here, right?

If you fix P to be one, you have to actually stop here, right? And plus, obviously, the random error. So, this is exactly what the AR1 model is. And again, the last component is kind of important, but we'll go into details slightly later, maybe in the later lectures, but one can actually prove when the corresponding AR1 model is stationary. Because again, remember the idea of stationarity or a time series being stationary is really important for forecasting, for modeling, etc.

And then you have a particular condition here. So, whenever the coefficient φ1 happens to be between minus 1 and 1, then we can simply say that the underlying AR1 model is stationary. Right. So, we can take up an example. So, let us say Yt equals 10 plus 0.4 Yt minus 1 plus et. Right.

$$Y_t = 10 + 0.4Y_{t-1} + e_t$$

Now, a couple of questions. So, this structure corresponds to which model first? So, again, clearly the structure corresponds to AR1 because we are stopping at 1 or p equals 1 here. Right. Or we are only taking one past value of y in that autoregression.

So, this is an AR1 model, and then the second question is, is this stationary? So, again, the answer is yes, because if you look at this coefficient, which is 0.4, right. So, 0.4 clearly lies between minus 1 and 1. So, this condition is checked for us. So, hence this time series structure or this time series model is stationary due to this single condition.

Now, probably we will discuss one extension of this, which is AR2. So, AR2 means what? So, again, just to mention, the P should be 2 here. And if you stop after taking two past values of the same series Yt and plus some random error, the model is called AR2. So, Yt equals C plus φ1 Yt minus 1 plus φ2 Yt minus 2 plus et.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$$

AR (2) model is stationary if:

$$-1 < \phi_2 < 1$$

$$\phi_1 + \phi_2 < 1$$

$$\phi_2 - \phi_1 < 1$$

And immediately here, you can actually see as to when the AR2 model is stationary or not stationary. The conditions are not easy here. So, conditions get pretty involved. So, by the way, any AR2 model is stationary if the coefficient Phi 2 is between minus 1 and 1, and Phi 1 plus Phi 2 should be less than 1, and Phi 2 minus Phi 1 is less than 1.

So, these conditions get pretty involved as you increase the order, which is P. So, one can actually imagine that if you go on for higher orders, let us say 3, 4, 5, 6, the conditions become so involved that one cannot prove this using pen and paper. So, one can actually make use of some hypothesis tests using some software and so on. But proving stationarity is kind of key. So, for some simpler models such as AR1 or AR2, this can be done. But for higher orders, one actually has to resort to some software.

Now, the second kind of the process is called moving average. So, moving average is denoted by MA(q), right? So, the order is termed as q here. So, what exactly is the difference between autoregressive and moving average? So, rather than using the past values of the forecast variable in a regression, an MA model or a moving average model uses the past errors in a regression-like model, okay. For example, if you look at this structure. So, how exactly is the structure written down? So, the corresponding time series Yt is a combination of the current error and a weighted average of all the past errors. So, this is more like a regression.

$$Y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

Consisting of all the errors and not the past values of Yt itself. So, in the autoregressive process, we had the past values of Yt itself, right? But here, what we have is the past values of the random error component, right? And again, the same kind of structure: if you stop at q, then we say that the order is q. So, this structure is a moving average with order q, or MA(q) model. So, a couple of points here: each value of Yt can be considered as a weighted moving average of past forecast errors. So, this is nothing but a weighted moving average of past forecast errors.

So, if you want to put forward a model for the irregular component, then pretty much a moving average model is suitable because moving average is nothing but a combination of some random errors, right? So, if you try to graph this, the graph would look like something completely random, okay? So, probably we'll end by discussing this ARMA model. So, ARMA is kind of an easy model to understand because ARMA means Auto Regressive Moving Average Process. So, this is nothing but a combination of the AR part and the MA part.

And naturally, there should be two orders or a pair of orders. So, ARMA (p, q), so what is ARMA? So, ARMA is nothing but a combination of AR(p) model and MA(q) model. And then, this is the structure. So, until this point, I will say that this is the AR structure, isn't it?

Because you are stopping at p. So, this is the AR(p) structure. And from this point onwards, this is the MA structure. So, MA(q). So, if you basically combine both these structures into a single equation, it is called the ARMA process or Auto Regressive Moving Average process. So, maybe in the next lecture, we will talk about some other moving average processes of, let us say, higher orders.

And we'll again focus on, let's say, some other kinds of time series processes which stem from this idea of autoregressive or moving average or the combination, which is ARMA, and so on and so forth, with some examples.

Thank you.