Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 01

Lecture 03: Stationarity in Time Series

Thank you. Hello all, welcome to this third lecture in this lecture series on time series modeling and forecasting using R. Just to give you a brief overview of what we covered in the last two lectures so that it'll be a nice flow to the course. So, as all of you know, we talked about, how exactly a time series data is different from any other statistical data that you might have seen earlier. We talked about several time series examples coming from different areas, be it finance or ecology or climatology, etc.

We also talked about the correct notation to follow when it comes to describing any time series observation. And if you remember, we talked about something like y and then in the subscript we wrote down t. So, that t stands for the time point. And towards the end of second lecture we also talked about choosing a particular time scale is really important and all of that kind of depends on the goal of the experiment. So, if you want to pick let us say daily data or if you want to analyze weekly data or monthly data or annual data that sort of depends on what the goal of the entire experiment is. we're into lecture three.

And initially we'll talk very briefly about few important functions that go inside any time series analysis. And I'm very sure that many of these functions would be kind of similar to any other statistical course that you must have seen. For example, you see a function which is called as a mean function. But the only difference here or I should say a subtle difference here, is that, as you see we have μ_t. So, this t stands for the time point again. So, rather than writing down simply μ which is a very well-known parameter to describe the mean, we kind of involve the subscript describing the time point also. And nothing changes majorly. So, what exactly is μ_t. So, μ_t is nothing but the expectation. So, this E stands for the expected value which is the usual notation to denote the mean or the average of any random variable.

$$\mu t = E[Y_t]$$

So, expected value of this time series Y_t. So, just to describe this briefly, the notation to denote a time series throughout this course will be Y_t. So, the first thing to note here is that from now on we will say that the time series would be denoted by Y_t where t is the underlying time frame. Thus, we can talk about all these functions later on. For example, the first one just to again summarize is the mean function which is nothing but the average of the entire process which is Y_t. And as written down here, so this μ_t kind of gives you the expected value of the process at time t. So, I think this point is again important that at a particular time point t, what exactly is the expected value of the entire process?

$$\sigma_t^2 = \gamma_0 = Var\ (Y_t) = E\ (Y_t - \mu_t)^2 = E\ (Y_t^2 - \mu_t^2) < \infty$$

$$\gamma_0 \geq 0$$

So, the first one is the mean function. Now, the second function that we will talk about is the variance function. So, you have done with the mean then the second is the usual variance. So, generally speaking we denote any variance function by sigma square. Now, again just to add that time frame in question we will write down sigma t square.

So, again where t stands for the time frame. But when you talk about a time series notation or time series literature, we have a slightly different notation, which is called as gamma. And then if you specify zero here, so it's called as gamma nought So, gamma nought stands for the variance function from a time series point of view. All right.

So, sigma t squared is nothing but equal to gamma nought in a slightly different notation, which is nothing but equal to variance of the time series process or time series random variable, which is Y_t. And again, this is nothing but the usual variance formula. So, expected value of the random variable itself minus its mean and then whole square. So, Y_t minus μ_t whole square and then you have an expectation outside. Now, again as many of you might already know that one can actually rearrange a few terms here and there and then get an alternative formula for the variance.

So, the expected value of Y_t minus μ_t whole square can be again written down as something like this formula,

$$\sigma_t^2 = \gamma_0 = Var\ (Y_t) = E\ (Y_t - \mu_t)^2 = E\ (Y_t^2 - \mu_t^2) < \infty$$

$$\gamma_0 \geq 0$$

So, expectation of Y_t square minus mean square. And again, throughout the course we require assumption that it should be finite of course. So, variance should be finite and on top of that since variance can't be negative, we have this another restriction which is a usual restriction. So, gamma nought or γ_0 has to be non-negative.

So, again the sigma t square in a slightly different notation we write it down as gamma 0 gives the variance of the process again at time t. Now, we will move on to couple of other functions. Now, on top you see the auto covariance function. So, again I am pretty sure that all of you must already be knowing the covariance function. So, covariance is nothing, but the dependency between any two random variables as compared to their means, right. So, the usual formula is something like this.

$$\gamma_{t,s} = Cov\ (Y_t, Y_s) = E\ [(Y_t - \mu_t)\ (Y_s - \mu_s)] = E(Y_t\ Y_s) - \mu_t\, \mu_s$$

So, expected value of a random variable. So, let us say Y_t in this case because we are talking about time series literature now minus its mean into a slightly different random variable Y_s minus its mean which is μ_s. The only difference to note here is that the notation firstly. So, notation is slightly different which is $\gamma_{t,s}$. Now again I will pause here for a minute and then just to remind all of you that what gamma 0 was.

So, again gamma 0 was the variance function whereas gamma t comma s describes the auto covariance function under any time series analysis. So, gamma t comma s is nothing but covariance between Y_t and Y_s which is again given by that formula. And again, one can actually simplify this formula to get an alternative structure which is expected value of Y_t into Y_s minus the product of their means. So, gamma t comma s is what? So, gamma t comma s gives the covariance between two time points.

So, let us say time t and time s for a particular time series which is Y_t. Now again, one last important point to note here is that you might be wondering that why it is called as auto covariance and then not just covariance function. So, you have very simple reasoning here. So, the terminology auto covariance is kind of used because remember that what is happening here, so we have a single time series which is Y_t, right? Because remember that what is happening here, so we have a single time series which is Y_t, right?

given at different time points. So, t could be 1, t could be 2, t could be 3, etc., right? So, what exactly do you mean by Y_t and Y_s? So, Y_t and Y_s are nothing but two different iterations of the same time series, right? So, by iterations I mean Y_t and Y_s are nothing but a single time series at two different time points which are t and s.

And essentially what you are doing is you are finding out the covariance between these two entities which come from a same series at different time points. So, it is more like you are finding out the covariance of a single series at two different time points. So, hence we write down auto covariance. So, auto means covariance of the same time series but at different time points.

So, unlike what we used to do in any other statistical courses. So, any other statistical courses what we do, is we kind of assume that you have two entirely different random variables. So, let's say X and Y. And then we talk about their covariance. So, covariance between X and Y, right?

But here the things are slightly different because Y_t and Y_s, even though they're two different random variables, they're coming from the same time series. So, all the properties or all the necessities that the random variable Y has or the structure that Y has is kind of preserved by both these entities, Y_t and Y_s, unlike here. So, hence the terminology is autocovariance and similarly autocorrelation. So, autocovariance or autocorrelation. So, the last function which is again important is the autocorrelation function.

$$\rho_{t,s} = Corr\ (Yt,\ Y_s) = \gamma_{t,s}\ /\ \sigma_t\ \sigma_s$$

Notation wise we use rho (ρ) the standard notation for any correlation function. Now, again the only difference is since we want to specify two different time points, we write down ρ and then t comma s. So, $\rho_{t,s}$ is nothing, but correlation between Y_t and Y_s and this is again I guess a standard formula. So, correlation is nothing, but covariance at the top divided by the product of the corresponding standard deviations. Ok. So, generally speaking the formula for correlation is, just to summarize quickly, is nothing but, covariance between x and y divided by sigma x into sigma y. So, this is a standard formula or standard definition of correlation between any two random variables x and y. Now again the only difference here is that since we are talking about the same time series but at different time points, we are going to call that as autocorrelation. Ok. So, just to summarize as to where we are at now. So, we discussed four functions. So, mean function, the variance function, the auto covariance function, and finally, the auto correlation function, right?

And the random variable under question in all these four functions is called as Y_t. So, I think now we slowly transition into a very important and very fundamental concept in time series analysis called a stationarity. So, we'll define what exactly you mean by stationarity in a while. But before that, just to give you a glimpse of or just a revision of some very basic ideas in statistics. So, the title of this slide is Join PDF of a Time Series.

So, by PDF, as many of you might know already, so PDF stands for Probability Distribution Function. So, how any statistical process is distributed. So, then you have all those distributions in place. So, let's say binomial or normal, exponential, gamma, Poisson, right? So, any distribution has its own PDF which is called as probability distribution function.

So, how are the values basically distributed on some scale? So, just to sort of revise some notation.

$$F_{X1}(x_1) = \text{marginal CDF}$$

$$f_{X1}(x_1) = \text{marginal PDF}$$

$$F_{X1, X2, \dots, Xn}(x_1, x_2, \dots, x_n) = \text{joint CDF}$$

$$f_{X1, X2, \dots, Xn}(x_1, x_2, \dots, x_n) = \text{joint PDF}$$

The moment we write down capital F, so capital F of any random variable X_1 for that matter is nothing but the CDF. So, CDF stands for cumulative density function or cumulative distribution function. So, just to sort of mention as to what this C stands for.

So, C stands for cumulative. So, cumulative density function or cumulative distribution function. All right. So, cumulative means what exactly? So, again, you want a definition of this capital F, which is nothing but probability that the random variable X_1 is less than equal to small x_1.

Right. So, we're more interested into the probability up to some point, which is X_1. So, sort of a cumulative in nature. So, hence it's called as cumulative density function. Right.

And on the other hand, if you write down a small f notation, so small f notation is the nothing but the PDF. okay, which is probability distribution function for a random variable x1. Now, again one last point to note here is that since we are only interested in a single random variable which is either x_1 here or x_1 here, hence we call this as a marginal CDF or marginal PDF, all right. So, again just remember that. So, if you are only interested in a single random variable, let us say x_1 here or x_1 there, then we will call that as a marginal CDF or a marginal PDF.

Now, on the other hand entirely, if you want a joined CDF or a joined PDF. So, what do you mean by that? So, again joined means you are talking about a distribution which can be fit on all these random variables together. So, x_1, x_2 up to x_n let us say. So, let us say if you have n random variables and then you want to quantify the overall distribution that can be put on all these n random variables.

It is called as a joined CDF or the joined PDF. So, I guess this slide is just to summarize the basic notations which you will see later on also in the course and again immediately in the next slide, but just a small refresher on what do you mean by marginal CDF, marginal PDF, join CDF and then join PDF. So, once you understand as to what do you mean by marginal CDF or join CDF or join PDF and all those things. we can tie all these things in a time series literature. So, for any observed time series at two different time points, let us say t and s, all right.

So, the marginal pdf of Y_t and Y_s can be written down as follows;

$$f_{Y_t}(y_t) \text{ and } f_{Y_s}(y_s)$$

$$f_{Y_t,Y_s}(y_t, y_s) \neq f_{Y_t}(y_t) * f_{Y_s}(y_s)$$

So, f of Y_t or f of Y_s, okay. So, again writing down the PDF or writing down the CDF is nothing different from any other stats course which is not time series in nature, all right. But now very important point is this that you see in the last equation. If you write down the joint PDF of a particular time series Y_t at two different time points t and s, which is nothing but given by this notation by the way.

So, small f yt comma ys that can't be written down as a product of the marginals. So, these are the marginals at those two-time points t and s. But the pressing question is that why do you see a not equal to sign here? So, generally speaking, let me backtrack the question and ask in a different way. So, when can you have an equality here? So, when can you say that the joint PDF would be equal to the product of the marginals?

When these two random variables are independent, right? But here again, remember from the first lecture or the second lecture that we had, that since you're talking about a time series literature, a time series terminology, Y_t and Y_s are not independent, right? We can't write down the joint PDF as the product of the marginals because Y_t and Y_s are dependent in nature and dependent on what again? So, dependent on this time frame t. So, again, just to visualize the entire idea of being dependent or being independent.

So, we can discuss a very simple example. So, you may remember the Google stock price example from lecture one or lecture two for that matter. So, take a very simple situation. So, stock price of Google today and then stock price of Google, let's say tomorrow. If you focus on the daily stock price of Google.

So, stock price of Google today and then stock price of Google tomorrow. And then due to some trend in the Google stock price or due to some repetition, due to some other pattern, one can actually hope that the price tomorrow is kind of dependent on the today's price in some sense. So, the price today is let's say $100 hypothetically. The price tomorrow cannot be $1000 suddenly. I mean, again, this can happen, but with a very, very small chance, right?

So, if it's 100 today, then you should expect that it might be, let's say, 102 or 105 or 98 or 95, is it? So, the Google stock price tomorrow is kind of dependent on its current price, right? So, when we talk about any such thing as to modeling the joint PDF under a time series scenario, we cannot put down an equality sign here because these two are not independent, okay? All right, so now I think we are entering the whole idea of stationarity and why is it important and so on and so forth. So, problems in handling the joint PDF when it comes to a time series terminology.

So, what exactly happens, so probably we'll read this first and then I'll try to explain briefly. So, since we have a single observation for each random variable Y_t at any given time point, inference is too complicated If the distributions or the moments, so moments mean what? So, moments mean the first moment is a mean, then slight variation of the second moment is a variance, etc. So, these are called as the moments.

So, inference is too complicated if distributions or moments change for all t or keep on changing for all t. Hence, it becomes very difficult to identify a joint distribution. So, again, remember what is happening. So, I will give you an illustration. So, let us say we try to plot the time series and then let us say on the x axis we have t as always. So, let us say 1, 2, 3, 4, 5 and so on.

And let us say the behavior of the time series is something like this. So, it goes up then probably comes down then stays here then again probably goes up then comes down etcetera alright. So, this is again a hypothetical example. So, let us say you have the time frame on the x axis this is the behavior of the particular time series which is Y_t ok. Now, if you take this point, so what exactly would be this point?

So, this point is nothing but Y_1, right? And then let us say if you talk about this point, so this point is Y_2 and so on, okay? Now again, just to reiterate what's written down here is that if you observe, as you keep on moving down the timeline, you only have one observation of the corresponding random variable Y_t at any given time point, isn't it? So, at time point 1, you have one sample observation, which is Y_1. At time point 2, you have again a single observation, which is Y_2, right?

At time point 4, you again have a single random variable observation, which is Y_4 and so on. If the process is changing rapidly at any given time point, so putting a single distribution on all the time points is difficult, isn't it? So, hence we require some simplification. Again, just to repeat one last time is that since any time series process keeps on changing at any given time point 1, 2, 3, 4, 5, 6 and so on. So, putting an overall distribution on this entire structure is kind of difficult.

So, hence we cannot handle the joint distribution very easily ok. So, way we require some simplification which is called as stationarity. So, we assume that or we kind of hope that a time series process should be stationary in nature. Now, again we will kind of elaborate this entire idea of stationarity in a bit more detail obviously, but this is the idea as to why we require stationarity all right. So, just to summarize stationarity in let's say a few bullet points.

So, what exactly is stationarity? So, stationarity is nothing but the most common assumption in any time series analysis. Now we assume that the probability laws governing the entire time series process do not change with time. Okay. Unlike what we saw in the previous example.

So, again, if you go back for a second, So, in this example, we clearly see that the probability laws or the distribution or the process is kind of changing rapidly at any given time point. So, we can actually say that such a process is not stationary. So, when can we say that a process is stationary? So, the probability laws governing the process do not change with time or the process is in statistical equilibrium.

So, statistical equilibrium means that it should be kind of smooth in one particular horizontal direction. So, you should not have any rapid movements or any rapid fluctuations in the time series. So, I think this is a brief idea about what do you mean by a stationary process? What do you mean by stationarity? So, if stationarity is present, we'll say that the time series or the underlying stochastic process is stationary.

Okay, so I think we now, we will talk briefly about two different kinds of stationary processes. Now, the first kind of stationary process is called a strict or strong stationarity. So, you have two alternative terms here. So, either strict stationary or strong stationary. Now, again what we will do here is that you consider a finite set of random variables.

So, let us say,

$$\{Y_{t1}, Y_{t2}, \ldots, \text{up to } Y_{tn}\}$$

Now, of course t1, t2 up to tn are different time points here. Okay. So, you have a stochastic process Y and then let us say that this is the sample from that time series or this is a sample this is an observed sample from the stochastic process Y. Now, I guess we discussed earlier also this is again a usual notation for the joint distribution function or the joint CDF because we have a capital F here this denotes the joint CDF.

So, again definition wise we should have that probability of each random variable which is Y_{t1} or Y_{t2} or Y_{t3} up to Y_{tn} should be less than or equal to their corresponding values. So, Y_1 or Y_2 up to Y_n right. So, I guess this is a standard definition of the joint CDF.

Okay, so let's say again, just to summarize, you have a stochastic process Y, this would be an observed sample from that stochastic process. And down here, this is nothing but the definition of the CDF or the joint CDF, okay?

Now we'll kind of see that what exactly happens to the joint CDF, okay? Okay, so now you actually have two different kinds of stationarity or two different kinds of stationary processes. Okay, so the first process is called as first order stationary. Now, we will kind of describe as to what you mean by this. So, we'll again go back here.

So, first order stationary and second order stationary are kind of different stationarity processes. So, now again just to describe what exactly is strong stationarity. So, in front of you, you can actually see different kinds of strong stationarity processes. So, we'll try to describe them one after the other. So, here as you see, we can talk about strong stationarity being of different orders.

So, the first one is called as the first order stationery. And obviously the second one would be called a second order stationery. And if you try to generalize that, then we get a process which is called as nth order stationery. So, I'll try to describe very briefly as to what you mean by each and every kind of stationery processes. By the way, all these kind of stationery processes come under the heading of strong stationery also.

So, initially a process is said to be first order stationary in distribution If it's one-dimensional distribution function is time invariant. Now again just to remind all of you that what do you mean by time invariant or time invariance. If you change the time span by a certain length. So, let's say if you fix some length which is K and if you extend that time span by that length K but still the overall joint distribution or the marginal distribution does not change. So, here mathematically one can actually write down as so capital F denotes the CDF.

$$F_{Y_{t1}}(y_1) = F_{Y_{t1+k}}(y_1) \text{ for any } t_1 \text{ and k}$$

So, CDF of the time series process Y at a certain time point t_1. So, let us say initially if you take t_1 is the same as the CDF of the same time series process Y, but at a slightly different time point which is t_1+k. So, the only consideration here is that if the underlying distribution or for that matter the underlying CDF is time invariant or they are exactly the same or they don't change, then we say that the process is first order stationary. And obviously this has to be true for any T1 and any K. So, one can actually take the jump which is K of any order or of any length. Now, similarly, if you try to extend the idea to, let us say, second order stationary.

So, what do you think a second order stationary process means?

$$F_{Y_{t1}, Y_{t2}}(y_1, y_2) = F_{Y_{t1+k}, Y_{t2+k}}(y_1, y_2) \text{ for any } t_1, t_2 \text{ and } k.$$

So, again, on the left-hand side, we have the join CDF now. So, since you are extending the first order to second order, we can bring in one more time series process, which is slightly different, which is, let us say, Y_{t2}. Okay. And then we can talk about the joint distribution of the same time series process Y, but at two different time points.

Let us say t_1 and t_2. All right. So, on the left-hand side, we have the joint CDF of Y_{t1} and then Y_{t2}. And again, the same idea. So, if the joint CDF is exactly same, if you shift the time points by some constant k, alright.

So, on the right-hand side what we have? So, on the right-hand side we have the CDF of y at time point t_1 plus k and at time point t_2 plus k, ok. Now, again this should hold true naturally for any combination t_1 t_2 and for any constant k. So, in a way, one can actually extend this to some other order. So, let's say third order stationary or fourth order stationary.

So, again, due to time considerations, what we can do now is we can generalize this entire thing and we can take it a step forward to describe what you call as the nth order stationarity or nth order stationary process. So, again, nothing changes here. So, the only difference is if you are taking more than one or more than two random variables at a time and talking about its joint distribution. So, the nth order stationarity means that on the left-hand side what we have is the join CDF of let us say n random variables. So, Y_{t1}, $Y_{t2,}$ $Y_{t3,}$ up to Y_{tn} and if that join CDF happens to be exactly equal to the join CDF of let us say Y_{t1+k}, Y_{t2+k}, Y_{t3+k} … up to $Y_{tn+k.}$ ok.

$$F_{Y_{t1},\ldots,Y_{tn}}(y1,\ldots,yn)=F_{Y_{t1+k},\ldots,Y_{tn+k}}(y1,\ldots,yn) \text{ for any } t1,\ldots,tn \text{ and } k.$$

So, probably it would be kind of easy to understand that either first order stationary or second order stationary or for that matter nth order stationary holds true if you are basically shifting the time points by some constant k. Now again naturally this should hold true for any such combinations of t_1, t_2 up to t_n and any constant k. So, probably in the next lecture what we will do is we will discuss the other kind of stationary process which is called as weak stationary. So, again, just a short while back, we described that you do have two different kinds of stationarity processes. So, strong stationary and then weak stationary.

So, just one point to make here is that strong stationary, as you must have seen by now, that strong stationary process takes into account the joint distribution in place. So, if you translate that joint distribution or if you shift the time stamp by certain order k, and even if the joint distribution stays the same, then we will say that the process is strong stationary. And then obviously the next kind of stationarity would be weak stationarity processes. So, by the name, so as the name suggests, so there'll be certain less restrictions when we talk about weak stationarity as compared to strong stationarity.

For example, let's say possibly a weak stationary process need not make any assumption about the distribution or any assumption about the underlying joint distribution or joint CDF for that matter and so on, all right? So, probably we will take up the idea of weak stationarity in the next lecture and then probably end this lecture here ok. Thank you.