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 08: Autocorrelation and the Partial Autocorrelation Functions

Thank you. Hello all. So, welcome to this new lecture in this course on time series forecasting with applications in R. Now, in today's lecture, we will again revisit some of the functions that we have seen earlier, precisely the ones that you see on the slides, which are nothing but the autocovariance function and the autocorrelation function. All right.

And then, probably from the last session, if you vaguely remember, we've discussed multiple time series models, such as, let's say, autoregressive and then moving average, or even the combination of AR and MA, which is called ARMA. So, probably in today's lecture and the one after this, we will kind of explore more on this front and then specifically write down some of the important functions of these two time series models, right? So, such as autocovariance and then autocorrelation, or for that matter, we will start with the mean function and then the variance function, and so on and so forth. So, just a disclaimer that both these sessions will be slightly mathematical in a sense that we'll try to involve slightly different notations, which you'll see on the next slide, of course, and then from there on, we'll kind of prove a few things also. So, it would be a systematic kind of a flow to how do you find a particular autocorrelation function or the autocorrelation function of, let's say,

Some of the basic models such as AR1 or MA1, right? Or for that matter, let's say ARMA1, and so on and so forth. And then, right towards the end, we'll bring in the flavor of a random walk also. And then we'll kind of explore a few things about random walk too. So, I think this slide is more of a revision kind of slide.

So, we will kind of revisit the autocovariance function and the autocorrelation function, alright. So, before we delve into the autocorrelation function, ah. So, probably all of you must have seen the above equation, which is gamma t comma s, right? Now, what exactly is gamma t comma s? If you recollect, gamma t comma s is nothing but the covariance between Yt and Ys, right. So, where t and s are particular time points, ok.

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

And again, as you remember, this is nothing but the formula for a general covariance. So, Yt minus its mean, which is μ t, into Ys minus its mean, which is μ s, and then you take the expectation of this product, right? And towards the end of this equation, this is the alternative structure of the same equation. So, the expectation of Yt into Ys minus the product of the corresponding means, all right. And what exactly is gamma t, s? It is nothing but the covariance between the time series values at two different time points, which are t and s, okay?

And similarly, if you vaguely remember the autocorrelation function. So, the autocorrelation function is nothing but the covariance in the numerator, which is gamma t comma s, and then divided by the corresponding standard deviations. So, σ t into σ s, okay? So, this rho t comma s gives the autocorrelation, while this gamma t comma s gives the autocovariance, okay? Now, just for the sake of it, we will kind of involve slightly different notations, and then we will see how things proceed, okay?

$$\rho_{t,s} = \mathbf{Corr}\ (\mathbf{Y_t}, \mathbf{Y_s}) = \frac{\gamma_{t,s}}{\sigma_t\ \sigma_s}$$

So, just to be in line with all the notations and probably some textbooks and so on and so forth. So, we introduce a slightly different set of notations. So, instead of saying gamma t comma s, what we will do is we will define the subscript as the lag between the observations. So, here we are using gamma k, right? And then, obviously, depending on the time point t, so gamma k of t. And this is nothing but the covariance between Yt and Yt plus k, all right.

$$\gamma_k\ (t) = Cov\ (Y_t,\ Y_{t+k}) = E[(Y_t - \mu_t)(Y_{t+k} - \mu_{t+k})] = E(Y_t\ Y_{t+k}) - \mu_t\ \mu_{t+k}$$

So, again, if you remember, how much is the lag between these two time points? So, the lag is nothing but k, right. So, the subscript we are using here in the autocovariance function is nothing but the lag between the two time series observations, which is k. Now, all the other things are kind of exactly similar to the previous slide. So, for example, the formula stays the same.

So, Yt minus μ t into Yt plus k minus μ t plus k, and then you take the expectation outside. And obviously, this is the alternative structure. So, the expectation of Yt into Yt plus k minus μ t into μ t plus k. Now, again, a small point to note here is that I can either take Yt plus k or Yt minus k, right? Because again, if you are only focusing on the lag between the two observations, right. So, again, the lag between Yt minus k and something like Yt is still k, right.

So, rather than writing down the covariance between Yt and Yt plus k, I can actually replace Yt plus k by Yt minus k. So, one can actually look at either side of Yt with a lag of k. So, again, gamma k gives you what? So, gamma k is nothing but the auto-covariance between the time series value Y at time points t and t plus k. And once you define the auto-covariance function, defining the auto-correlation function is not difficult at all. So, again, notation-wise, what do you have?

$$\rho_k = Corr(Y_t, Y_{t+k}) = \frac{\gamma k}{\sigma_t \sigma_{t+k}}$$

So, you have rho k, which is nothing but the correlation between Yt and Yt plus k, which is, as per the formula, nothing but gamma k. So, gamma k is the covariance divided by the standard deviations. So, σ t into σ t plus k. So, hopefully, if you stick to this notation, probably this set of slides and the next set of slides would be kind of very easy to follow, all right. So, again, just to summarize, rather than writing down gamma t comma s, something like that, we are using a slightly different notation, which is nothing but gamma k, and what exactly do you mean by k? So, k is nothing but the lag between the observations Yt and Yt plus k, all right.

Okay, so now we can talk about some very elementary properties of the auto-correlation function, and probably all these are kind of standard properties. So, as mentioned earlier, if you have any stationary process, right? So, again, remember, what do you mean by a stationary process? So, let's say you have a constant mean, and you have a constant variance, and the auto-correlation function or the auto-covariance function should only depend on the lag, right? So, in that case, what happens is rho k is exactly equal to rho of minus k.

And this is again coming from the point that we discussed a short while back. For example, if you write down the correlation between something like Yt and then Yt plus k, this should obviously be equal to the correlation between, let us say, Yt and then Yt minus k. So, rho k equals rho of minus k, and obviously, the second property is kind of a pretty standard property when it comes to any correlation. So, the absolute value of rho k should be less than or equal to 1. Or in other words, rho k should lie between minus 1 and 1.

And the third property is kind of interesting. So, it's called the non-uniqueness property. So, what do you mean by the non-uniqueness property? So, let's say if you have a stationary normal process. So, a stationary normal process means the distribution or the underlying distribution is normal.

And then probably many of you might already know that if you have a normal distribution. So, any normal distribution is completely characterized by its first two moments, which are nothing but the mean and the variance, all right. So, similarly here, if you start with any stationary normal process, that process is completely determined by its mean, variance, and autocorrelation function. However, if you have a non-normal process, which is not a normally distributed process, then one can actually find several such non-normal processes with the exact same ACF or the exact same autocorrelation structure. So, just to summarize this third point.

So, let us say I can actually bring to the table multiple such non-normal time series processes having the exact same autocorrelation function. So, in a way, this autocorrelation function is not unique to a particular normal process or a particular non-normal process. Make sense? So, these are some standard properties when it comes to, let us say, defining the autocorrelation structure or where exactly should an autocorrelation lie. So, between minus 1 and 1, and finally, the non-uniqueness of the autocorrelation function.

Now, probably what we will do is we will try to study the ACF function of some of the basic time series models, right. So, I think in the last class we kind of discussed white noise, right. So, if you vaguely remember, what exactly is white noise? So, white noise is nothing but a bunch of random errors, right. So, if you want to plot white noise, it will be nothing but a combination of random fluctuations, right.

So, something like that. And obviously, the assumptions are that there should be a constant mean and constant variance. So, even if you see here, the variance is also kind of constant along the entire timeline, and then you have this constant mean, which is μ, all right. So, such a process is called a white noise process, and on top of that, if you identify or if you kind of put a distribution on this white noise, which is nothing but the normal distribution. Then the name is Gaussian white noise.

So, Gaussian is nothing but a normal distribution, right? So, Gaussian distribution is a different name for normal distribution. So, in a sense, Gaussian white noise means what? So, Gaussian white noise is nothing but a standard white noise along with the assumption that the underlying distribution is a normal distribution, okay? So, in a way, if you notice, we don't have many things on this slide.

So, this function ρk, gives you nothing but the autocorrelation function of this purely random process or a Gaussian white noise process. So, if you look carefully here, ρk is structured as follows. So, ρk is exactly equal to 1 if k is 0. Now, one point to note here is that if k is 0, what do you mean by that?

So, if k is 0, you get something like correlation between Yt and Yt. And all of you know that the correlation of any random variable with itself has to be 1 always. So, rho k has to be 1 if k is 0, and for any other lag, let us say 1, 2, 3, or 10, the correlation should be 0. Does this make sense? It should make sense because again, going by the properties of white noise.

So, what exactly are the properties of white noise? So, in white noise, the mean should be constant, and white noise should be IID. So, all the observations in white noise are independent, and hence the correlation has to be 0, basically, okay. So, this is a very simplistic-looking structure of the ACF of a purely random process or a Gaussian white noise process, okay. All right.

So, we will now shift our attention slightly to a sample ACF. So, remember one thing: obviously, in all practical situations, the underlying population is unknown, right? So, if you are sitting in any statistics course or, for that matter, a time series course, etc., what we assume is that all the population parameters are kind of unknown, and then what we do is take a sample to estimate all the population parameters. So, similarly here, let us say if you observe this data. So, Y1, Y2, up to Yn be the observed time series, right?

$$r_1 = [\Sigma_{t=1}^{n-1} (y_1 - \bar{y}) (y_{t+1} - \bar{y})] / [\Sigma_{t=1}^{n} (y_t - \bar{y})^2]$$

Now, again, one can actually think of many examples. So, let us say Y1, Y2, Yn would be, or they might be, let us say, daily stock prices of Google, you know, or let us say daily temperatures at a location, anything right, but essentially Y1, Y2, up to Yn are nothing but the observed time series. Then one can actually write down the sample ACF. So, sample ACF at lag 1, right? So, a couple of important points here.

So, this is the sample version of the correlation function, firstly, and at what lag. So, we are writing down at lag 1, OK, and then this is the formula. Now, again, if you see this formula for a second, the formula is not very difficult to digest. So, what we have done is we have kind of replaced all the means by the sample means. For example, Y bar here or Y bar there, right, or essentially Y bar here, right.

So, this entire numerator is nothing but the covariance, is it not, because summation Yt minus Y bar into Yt plus 1 minus Y bar. So, the numerator is nothing but the covariance, and then the denominator is nothing but the variance, right? Because remember one thing that, let us say, if you assume that you have a stationary series, so both the means are the same, Y bar and Y bar, so the mean is constant throughout any lag. And we can also assume that the variance is also constant. So, in the denominator, if you take something like under root of variance of Yt and variance of Yt plus 1, but since both the variances are the same, we can actually write down something like that. Make sense?So, this is a classical formula of a sample ACF at lag 1. So, r1 stands for sample ACF at lag 1, and similarly, one can kind of very easily extend this to any other lag.

So, let us say sample ACF at lag k would be just a small extension. So, rk, and notice where exactly you see the difference. So, you see the difference probably here because now we want to find out the sample ACF at lag k directly.

$$r_k = [\Sigma_{t=1}^{n-k} (y_t - \bar{y})\ (y_{t+k} - \bar{y})] \ / \ [\Sigma_{t=1}^{n} (y_t - \bar{y})\ ^2] = c_k \ / \ c_0$$

So, instead of Yt plus 1, we have replaced that by Yt plus k, essentially, alright. So, the formula becomes summation Yt minus Y bar into Yt plus k minus Y bar divided by the overall variance, right. And just to be in line with all the notations used in the textbook and so on. So, this could be kind of determined by CK by C naught. So, this is a kind of a special notation to denote rk.

So, Ck is the numerator, by the way, and then C nought is the denominator. So, we have actually discussed two things. So, sample ACF at particular lags, right. So, at lag 1, which is r1, and then sample ACF at lag k, which is nothing but rk. Alright, so now we will pay attention to a particular very important plot in time series literature, which is called the correlogram or the ACF plot.

So, what exactly do you mean by a correlogram or ACF plot? So, the idea is kind of very simple. So, if you try to graph rk against k. So, remember, what is rk? So, rk is nothing but the ACF function or the autocorrelation function at lag k. So, if you try to plot rk against k, such a graph is called either a correlogram or the autocorrelation function plot. And why exactly is it important?

Because any ACF plot plays a very important role in model identification. So just to give you some ideas, let's say how exactly such a plot would look. So, let us say on the x-axis you will have k. So, k is the lag, right, and then on the y-axis you will actually plot the corresponding correlations. So, k versus rk, right, and upon looking at the structure of such a plot, which is nothing but the correlogram or the ACF plot, one can actually identify the underlying time series model easily. OK. So, we will do a small exercise slightly later on this front.

As to how do you kind of identify the model just by looking at the plot and so on. But the important part in this slide is to understand the idea of a correlogram, OK. So, the correlogram is nothing but a simple graph of k versus rk, right. So, now, the other side of the coin is a slightly different autocorrelation function, and probably I am kind of introducing this idea for the very first time, and the name of such an autocorrelation function is called the partial autocorrelation or PACF, alright. So, the partial autocorrelation function.

Now, again, the idea is kind of easy if you follow this twice or thrice, right, but what exactly is PACF? So, let us say PACF of any order k, let us say alpha k. So, alpha k denotes the PACF. It is nothing but the partial correlation coefficient between Yt and Yt minus k, right. So, this is nothing but kind of the correlation between Yt and Yt minus k, but you have something extra here. So, conditional on the intermediate values of the process.

So, what do you mean by that? So, let us say this is this kind of a conditional correlation. So, how do you write down conditional probability or conditional correlation by a vertical line, something like that, right? So, again, if you read the statement. So, the correlation between two variables Yt and Yt minus k conditioned on all the intermediate values.

So, what exactly are the intermediate values or what values fall between these two? So, which are nothing but let us say Y_{t-1}, then Y_{t-2}, etcetera, up to Yt minus k plus 1, something like that, right. So, this formula is nothing but the formula for the PACF or the partial autocorrelation function, ok. Or in other words, this PACF or alpha k is nothing but the autocorrelation between Yt and Yt minus k, removing their linear dependence on the intermediate values, which are nothing but Y_{t-1}, Y_{t-2}, Y_{t-3}, and so on and so forth, up to Yt minus k plus 1, ok. So, even if you understand the simple idea that, in practice, we have two very important autocorrelation functions.

So, one is the ACF function, and the other one is the PACF function, right? And for that matter, one actually has plots for both these. So, one can actually construct plots for both the ACF function and the PACF function. And collectively, these two plots are very important in model identification. So, if you want to find out that your practical data should follow what kind of a model.

So, the very first thing that researchers do is plot these two plots, which are nothing but the ACF plot and the PACF plot. And then, using the structure of these two plots, they are able to identify the underlying model. All right. Now, what we will do is expand on some of the elementary time series processes and then try to find out the ACF functions for those. So, the first one is the autoregressive process.

So, probably all of you know the structure. Now, again, just for revision, I have written down the structure here. So, if you recall, the general structure of an autoregressive process of order P is something like that. So, Yt is regressed on its historical data. So, Y_{t-1}, Y_{t-2} up to Yt minus P, and then eventually you have an error term, which is et.

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

And how exactly do you identify the order? So, the number of historical terms of Y that you take determines the order. So, for example, here, if you are taking P such historical terms, then the order becomes P, okay. So, this is the general structure of the ARP process, and some properties about all the parameters are as follows. So, et is nothing but an IID sequence with mean 0.

And a fixed variance, which is σ square e, and all these are the model parameters. So, φ 1, φ 2, φ 3 up to φ P, and then C, and then σ square e, all right. So, I think this slide just gives you a revision of a general autoregressive process of order P. Now, further, if you assume that all the means or the mean throughout is a constant, which is mu. So, the expectation of Yt is the same as the expectation of Y_{t-1}, which is the same as the expectation of Y_{t-2}, and so on and so forth, up to the expectation of Yt minus p, which is nothing but some constant mean, which is mu, okay.

So, the first line just tells you that you are kind of assuming that you have a constant mean throughout. So now, can somebody tell me what would happen if you take expectations on both sides of this ARP model? So again, if you go back for a second. This is what you have, right? So, looking at the structure here, if I take expectations on both sides, what would you have?

So, the expectation of Yt is nothing but c, and c is a constant, right? And then plus φ one times the expectation of y t minus one, right? So, if you follow, and then dot, dot, dot, plus the last term is φ p times the expectation of Yt minus p, right? And now, the last term would be 0 because the expectation of the error is 0. Hopefully, this should be clear now.

Alright. So, if you take expectations on both sides, this is exactly what we get. Now, as per our assumption that all the means should be constant. So, what do you have? So, this is nothing but μ equals C plus φ 1 times μ plus φ 2 times μ up to φ P times mu.

So, I am basically replacing all the expectations by μ because we are assuming that the mean is constant. Now, can't you solve this equation for μ and then get a formula? So, of course, we can. And this is exactly what is given in the next slide. So, if you take expectations on both sides, this is the first step.

And then, bringing all the μ's on one side, this is nothing but the value of that expectation. So, C divided by 1 minus φ_1 minus φ_2 minus φ_3 dot dot dot minus φ_p. So, again, just to summarize, if you assume a constant mean throughout, this is nothing but the value of that mean. Now, what we will do is we will consider some very basic processes. So, let us say AR (1).

So, what do you mean by AR (1)? So, AR (1) means the order is 1. So, you are regressing the current value on just one past value, which is Y_{t-1}. So, this is the structure of AR (1). So, Y_t equals some constant c plus φ_1 multiplied by Y_{t-1} plus the error term, OK?

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

Now, a very interesting property of an AR (1) process is as follows. So, given Y_{t-1}, Y_t actually becomes independent of all the previous values, is it not? Because this equation does not contain any of the previous values. And then, given the fact that, let us say, Y_{t-1} is a constant. So, given Y_{t-1} means what? So, given Y_{t-1} means that one can assume that this Y_{t-1} is a constant or it behaves as a constant.

So, assuming that Y_{t-1} behaves like a constant, Y_t becomes independent of all the previous values. So, Y_{t-2}, Y_{t-3}, etc. And such a property is called the Markovian property. And hence, this AR (1) process is, in fact, a Markov process. So, just to summarize what exactly you mean by a Markov process.

So, the Markov process means that the current value is only dependent on one past value. So, for example, here, since Yt only depends on Y_{t-1}, obviously given the fact that Y_{t-1} act as a constant. So, an AR1 process can be thought of as a Markov process. Now, again, finding the mean is not difficult. So, probably you can pause the video if you want and then try finding the mean of this AR1 process.

So, one easy way to do it is if you go back, this is the general mean for ARP, right. So, this is the general mean for ARP. So, simply, if you replace P to be 1 here, you will get the mean of an AR1 process, right, and which is exactly what you see on the next slide. So, the mean of the AR1 process happens to be mu, which is C divided by 1 minus φ 1, right. All right.

So, now, further, So, what we can assume is we can still assume that since we are assuming all the means to be constant. So, why not just put some constant value which is 0 for a trivial purpose. All right. So, assuming that C equals 0, which implies μ to be 0.

So, the constant mean, in fact, is nothing but 0. And we are putting one more assumption here that the process is, in fact, variance stationary. So, what do you mean by variance stationary? So, variance stationary means that all the variances should also be the same and constant. So, for example, gamma 0 is a notation for variance, right.

So, probably we have seen this notation in the earlier lectures. So, gamma 0 is nothing but variance of Yt equals variance of Y_{t-1} equals variance of Y_{t-2} and so on up to the last term. And we are also assuming that all the variances are the same and equal to this guy which is σ square y. Now, obviously σ square y does not depend on t. So, variance in fact does not depend on t and all the variances at any lags are equal, ok. So, if you now combine these two assumptions that the overall mean is 0 and all the variances are equal and equal to σ square y. So, can you essentially find out the variance of the AR 1 process, ok.

$$\gamma_0 = V(Y_t) = V(Y_{t-1}) = \dots = \sigma^2_y$$

$$\sigma^2_y = E\,[\varphi_1\, Y_{t-1} + e_t]^{\,2} = \varphi^2_1\, E(Y^2_{t-1}) + E(e^2_t) + 2\varphi_1\, E\,(Y_{t-1}\, e_t) = \varphi^2_1\,(\sigma^2_y + \sigma^2_e)$$

Now, again going back. So, this is the AR 1 process structure which is given here, right. So, if you kind of write down variances on both sides now. So, rather than expectation if you write down variances on both sides then what will you have? And this is kind of explained on the next slide.

So, σ square y would be nothing, but the expectation of the AR 1 structure. So, this is nothing, but the AR 1 structure and the square of that because remember one thing that variance of Yt is nothing, but the expectation of Yt square minus the expectation of Yt whole square, isn't it? So, this is standard formula for variance. So, variance of any random variable is nothing but expectation of that random variable square minus the whole square of the expectation of the random variable. So, using this structure we have something like that.

So, variance of Yt equals E of Yt square minus E of Yt whole square. Now, since we are assuming that the expectation is 0, so this part goes away by the way and we are only left with variance of Yt being expectation of Yt square which you see here. So, σ square y is nothing but expectation of φ 1 Y_{t-1} plus et whole square and now I can actually expand this. So, if you open the brackets what will you have? So, you will have something like φ 1 square and then expectation of Y_{t-1} square plus expectation of tt square plus the cross-product term which is 2 into φ 1 into expectation of Y_{t-1} into et. Now, the most important question is why is this cross-product term equal to 0?

So, if you visualize this for a second, what we have is φ 1 is a constant, firstly, right. So, φ 1 comes out of the expectation, and then inside the expectation, you have Y_{t-1} into et, right. Now, again, if you go back for a second and look at this general AR 1 process. So, what exactly would be Y_{t-1}? So, Y_{t-1} would be nothing but, similarly, c plus φ 1 y $_{t-2}$ plus e_{t-1}, right, and so on.

So, similarly, I can write down Y_{t-2}, Y_{t-3}, Y_{t-4} etcetera. Now, if you see this Y_{t-1} structure, Y_{t-1} does not depend on et, right; it only depends on e_{t-1} and the further lags. So, e_{t-1}, e_{t-2}, e_{t-3}, etcetera, but not on et, right. So, I can write down a small assumption here on the next slide that Y_{t-1} is kind of independent. So, Y_{t-1} is independent of et, and hence the expectation has to be 0, right.

So, since Y_{t-1} and et are independent, the expectation of their product has to be 0, ok. And then, so this entire cross product term becomes 0, right, and then this boils down to φ 1 square, and then the expectation of Yt minus square is nothing but the variance as per this assumption, ok. So, if you remember, we are assuming constant variances, right, and then the expectation of e t square would be nothing but the variance because the expectation of e t is 0. Again, so finally, it boils down to this, all right. And then, if you combine the LHS, LHS is σ square y, and RHS is this. So now, I can actually solve for the variance, isn't it?

So, this implies that my σ square y would be nothing but σ square e divided by 1 minus φ 1 square. Ok. So, essentially, the variance of any AR 1 process, assuming these two properties hold true, is nothing but given by that formula. So, σ^2e divided by 1 minus φ 1 square, ok. So, probably now, what we will do is we will kind of explore more on a couple of other properties of AR 1 and then slowly transition into the moving average process, right.

So, we will kind of deal with these two sides of the coin: auto-regressive and then moving average. And then, what we are trying to essentially do is kind of simplify a few properties of both these processes. For example, let us say auto-covariance, then auto-correlation, and then, of course, under some assumptions. So, if you assume, let us say, stationarity. So, stationarity means constant mean or constant variance.

Then, I think, we will kind of extend this to the next lecture also. Thank you.