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 09: ACF and PACF for Some Time Series Processes

Hello all. So, welcome back to this new lecture in the course on time series forecasting with applications in R. So, again, just to give you a very quick refresher of where we stopped in the last lecture. So, if you remember, we were discussing some properties of the AR1 process or the autoregressive process of order 1. And up to now, we found under some assumptions, for example, constant mean and constant variance.

So, essentially assuming that the process is stationary, we kind of derived the variance of the AR1 process, right? So, this is nothing but the variance of the AR1 process. So, σ square e divided by 1 minus φ 1 square, right? Now, again, in today's lecture, we kind of extend more upon such ideas. So, we will try to derive the ACS function of the AR1 process and then slowly transition into moving average processes and probably random walk processes, okay?

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

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

$$= \sigma^2_y = \frac{\sigma^2_e}{(1-\varphi^2_1)}$$

So, now how would you derive the autocorrelation function of a particular AR1 process? Now, again, remember that why are we discussing such simpler processes is that because one can actually derive, let us say, the mean or the variance function or the ACF function of such simpler processes using a pen and paper. Otherwise, one cannot derive even the mean function or the variance function using a pen and paper. So, one requires some software for that. But probably all these exercises would be kind of useful to understand how to play around with, let us say, lags or how to kind of play around with the assumptions and so on and so forth.

Now, again, assuming that the mean is constant and equal to 0, as before. Now, γ_1. So, γ_1 means what? So, γ_1 is nothing but the covariance at lag 1. So, it would be something like the expectation of yt into Y_{t-1}.

$$\gamma_1 = E\,(Y_t\, Y_{t-1}) = E[(\varphi_1\, Y_{t-1} + e_t)\, Y_{t-1}] = \varphi_1\, Y_0$$

Now, of course, I will pause here for a second and then tell you that there should be one more term. I mean, the covariance term is what? So, covariance is the expectation of yt and then Y_{t-1} minus the expectation of yt into the expectation of Y_{t-1}, right? So, this is the exact formula for covariance, but under the assumption that both these means are 0, right? So, both these means are 0.

So, covariance at lag 1, which is γ_1, essentially reduces to that thing, ok? So, now what we can do is we can input the model here. So, yt is what? So, yt has an AR 1 structure, which is nothing but φ 1 into Y_{t-1} plus e t into Y_{t-1}. Now, notice here that this is a product that you have, right?

So, the expectation of, let us say, φ 1 Y_{t-1} into Y_{t-1} would be nothing but φ 1 into γ 0 because the subscripts are the same. So, if the subscripts are the same, the expectation of Y_{t-1} into Y_{t-1} is nothing but the variance, which is γ 0, and then you have this coefficient outside, which is φ 1, and essentially, as we discussed also in the last lecture. So, this e t and Y_{t-1} are independent. So, the expectation of e t into Y_{t-1} is nothing but 0. Basically, and why is that? Because we will try to explain this.

So, the expectation of e t into Y_{t-1}, since they are independent, I can break it down into the product of individual expectations, right? So, the expectation of e t into the expectation of Y_{t-1}, and as per the assumption. So, this is equal to 0, and again, this is also equal to 0 as per the assumption. So, the expectation of their product is automatically 0, ok? So, further, the expectation of y(t-k) into e t is also 0, and again, using the same idea that y(t-k) does not depend on e t, right?

$$E(Y_{t-k}\, e_t) = 0$$

So, probably y(t-k) depends on all the previous error terms before t - k, right? So, y(t-k) is completely independent of e t. So, their expectation can be again broken down into individual expectations and their product, and both of them have to be 0s. Now, again, why do we require this assumption? So, you will slowly see which is coming in the next equation. So, similarly, I can actually derive γ_k. So, γ_k means what?

So, γ_k is nothing but the autocovariance at lag k, which is nothing but the expectation of yt into y(t-k). Now, again, if you replace the AR1 structure in place of yt, which is nothing but that. So, φ 1 into Y_{t-1} plus e t multiplied by y(t-k). Now again, notice one thing here: if you open the brackets and then take the product here, it will be the expectation of φ 1 into Y_{t-1} into y(t-k). So, φ 1 comes outside, and the expectation of Y_{t-1} into y(t-k), as per the notation, is nothing but γ_k minus 1. Because if you notice the lag, the lag between t-1 and t - k is nothing but k minus 1.

$$\gamma_k = E\,(Y_t\, Y_{t-k}) = E[(\varphi_1\, Y_{t-1} + e_t)\, Y_{t-k}] = \varphi_1\, Y_{k-1}$$

$$= \varphi_1\, \varphi_1\, \gamma_{k-2} = \ldots\, \varphi_1^{\,k}\, \gamma_0 = \varphi_1^{\,k}\, \frac{\sigma e^2}{(1-\varphi_1^2)}$$

So, I can actually write down γ_k minus 1 into φ 1. And using the above property, the expectation of e t into y(t-k) happens to be 0. Now, if you notice this, something interesting should happen here. So, I can actually do this recursively. So, how exactly recursively?

So, φ 1 into γ_k minus 1. Now, what I can do is I can expand γ_k minus 1. So, γ_k minus 1 will be nothing but a function of γ_k minus 2. And similarly for all the future lags. So, γ_k minus 2 would be a function of γ_k minus 3 and so on.

So, essentially, I can write down this is nothing but φ 1 into φ 1 into γ_k minus 2, which is nothing but φ 1 into φ 1 into φ 1 into γ_k minus 3, etcetera. And the last term would be something like φ 1 to the power k into γ nought, okay. So, probably, if you are not very confident, the small suggestion is you can actually try this out, right. So, the next thing you have to do here is write down a similar function for γ_k minus 1 and probably γ_k minus 2, γ_k minus 3, etcetera, until you get this recursive kind of structure. So, essentially speaking, this is nothing but φ 1 to the k, and now I can replace γ 0.

So, γ 0 is nothing but the variance of AR 1. So, we can use the same variance structure from the last class and replace it here. So, we kind of get that γ_k for any AR1 process is nothing but φ 1 to the power k multiplied by σ square e divided by 1 minus φ 1 square. And now, once we get hold of γ_k, I can easily derive ρ k, which is nothing but the ACF. So, ACF is nothing but γ_k divided by γ nought, which reduces to only φ 1 to the power k.

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1^k$$

Okay. So, just to summarize what we have from this entire slide, we kind of manipulated a few assumptions here and there. So, let us say, assuming that the mean is constant and variances are constant and so on and so forth, we kind of arrive at this conclusion that the autocorrelation function of a particular AR1 process is nothing but given by φ 1 to the power k. All right. So, now what we have here is we have a simulated kind of experiment, so we are kind of simulating from this model which is AR1, and the coefficient is 0.6, right? So, I think this slide just tells you the path of that time series process. So, this is exactly how the AR1 process would look like, as a simulation, of course.

Where that φ 1 happens to be, so this 0.6 is nothing but φ 1, the value of φ 1. So, the coefficient is 0.6. Okay. But now, on this slide, you will see how the ACF function of the same model looks like. So, the ACF function of a stationary AR1 process where φ 1 equals 0.6, if you remember, the formula for ACF is nothing but φ 1 to the power k, where k is the lag, right? We derived it just a short while back. So, if you look at this plot, it starts at lag 1.

So, let us say lag 1 is here somewhere. So, at lag 1, which means k is 1, the ACF should be φ 1, right? And φ 1 is nothing but 0.6. So, this is nothing but 0.6 here. So, now if you move down the value of k. So, let us say k is 2.

So, if k is 2, the value is nothing but φ 1 square. So, φ 1 square is nothing but 0.6 square, which is nothing but 0.36. So, at lag 2, if you see, the value is close to 0.36, and so on. So, by the way, this plot is nothing but a correlogram or an ACF plot of an AR1 process where the coefficient φ 1 is nothing but equal to 0.6, okay. All right.

So, now, we will kind of transition to the MA processes. So, so far, we have kind of derived some properties of the autoregressive process. So, now, we will kind of derive the same properties for some of the basic moving average processes, right. So, again, if you recall, the general structure of an MAQ process is something like that. So, what exactly is the difference here again?

So, the difference is yt is regressed on the past errors instead of the past values of yt itself. So, yt happens to be nothing but c, which is nothing but the overall intercept or the constant mean, and then plus all the errors. So, e t plus θ 1 e_{t-1} plus θ 2 e_{t-2}, etcetera, up to θ q e t - q. Right. And again, the same assumption.

$$\boldsymbol{Y_t = c + e_t + \theta_1\, e_{t-1} + \theta_2\, e_{t-2} + \ldots + \theta_q\, e_{t-q}}$$

So, what exactly are et's? So, et's are nothing but random errors. So, IID sequence with mean 0 and variance σ square e. So, essentially all these are the underlying parameters. Right. So, again like I said, this slide is more like a revision slide just to recall the idea of a particular MAQ process.

Okay. Now, again further we kind of do a similar thing as we did for the AR processes also. So, let us say upon taking expectations on both sides of the MAQ process, right. So, let us go back for a second. So, this is the MAQ process.

So, what would happen if you take expectations on both sides? So, expectation of yt happens to be C. So, C is a constant plus expected value of et. Now, again all the θs are constants. So, you can bring them out. So, θ 1 into expectation of e_{t-1} and then last term would be θ q and then expectation of e t - q, right. So, this implies that expectation of yt would be nothing but C, right, and all these expectations are kind of zeros because they are expectations of the error terms.

So, as per the assumption, errors are nothing but random errors with mean 0, ok. So, essentially speaking, what we have is something like this. So, expectation of yt is nothing but a constant mean which is nothing but the overall intercept that we have in the model which is C, ok. And similarly, what we can do is we can again take variances on both sides. So, what would happen?

So, on the left-hand side, we will have the variance of yt, then the variance of c is 0 because c is a constant, right. So, the variance of any constant is 0, and then you have the actual model. So, the variance of e t, then in the model of MAQ, we have θ 1. So, if you bring out θ 1 of a variance, you have to square that. So, it becomes θ 1 squared variance of e_{t-1} and so on up to θ q squared, and then the variance of e t - q, right.

$$\boldsymbol{V(Y_t) = V(e_t) + \theta_1^2 V(e_{t-1}) + \cdots + \theta_q^2 V\left(e_{t-q}\right) = \sigma_e^2 \sum_{j=0}^{q} \theta_j^2}$$

And now what we can do is we can actually combine all these ideas, right. So, by the way, all these variances are kind of variances of error terms, and again, as per the assumptions, all of them are equal and equal to σ square e, right. So, essentially, I can bring out σ square e from each of these individual terms. So, σ square e from here, from here, from there, etc. So, overall, if you see, we have a σ square e term outside and a summation of the squared coefficients inside.

So, θ j squared, j going from 0 to q. And then one small assumption is required here that θ 0 squared is nothing but 1, basically. Because j starts from 0, so θ 0 squared is assumed to be 1. And then we will have θ 1 squared, θ 2 squared, θ 3 squared up to θ q squared. Hopefully, this is clear.

So, the overall mean of any MAQ process is C, and the overall variance of any MAQ process happens to be this thing. So, σ square e into the summation of θ j squared, j going from 0 to Q. Now again, as discussed, a short while back, we will kind of involve some basic processes. So, we will start with the MA 1 process, right. So, the structure of the MA 1 process is as you see here.

$$Y_t = c + e_t + \theta_1 e_{t-1}$$

So, yt equals c plus e t plus θ 1 e(t-1), ok. Now again, as per the calculations we have done a short while back. So, the overall mean function is c, while the variance function becomes σ square y, as per the notation this is equal to γ 0. Which is nothing but σ square e into 1 plus θ 1 square, right? Now, how about the autocovariance function and then eventually the ACF function of an MA(1) process, right?

$$\sigma_j^2 = y^0 = \sigma_t^2\left(1 + \theta_1^2\right)$$

So, for that, what we can do is we can start by defining at smaller lags. So, let us say γ_1. So, γ_1 is what? So, γ_1 denotes the covariance between yt and Y_{t-1}, right? So, here I can actually replace the actual MA(1) model, right.

$$\gamma_1 = Cov(Y_t, Y_{t-1}) = Cov[(c + e_t + \theta_1 e_{t-1})(c + e_{t-1} + \theta_1 e_{t-2})] = \theta_1 \sigma^2_e$$

So, this is nothing but Yt, and then this is nothing but Y(t-1), right? So, from the previous slide, if you go back for a second, this is the MA(1) structure, right. So, I can actually write down Y_{t-1} also, right. So, Y_{t-1} would be what? So, Y_{t-1} would be C plus e(t-1) plus θ 1 e(t-2).

So, similarly, you can actually replace yt and Y_{t-1} here. So, we will have covariance of the product of yt and Y_{t-1}, or essentially it should be covariance of this,that. So, yt,Y_{t-1}. Now, finding covariances of such a structure is kind of very easy. So, what you do is you spot for similar subscripts. So, wherever if you see similar subscripts, for example, t-1 here and then t-1 there, then the covariance essentially reduces to variance, right?

And for all the other terms where the subscripts are not the same, the covariance would be 0 because essentially you are finding out the covariance between error terms where the subscripts are not the same. So, according to the assumption, So, all the errors are IID, right? So, whenever the subscripts are not the same, for example, the covariance between e t and e_{t-2}. So, these are completely different error terms.

So, as per the IID assumption of error terms. So, errors are independent. So, such covariances happen to be 0. So, essentially you only have to focus on such subscripts which are kind of the same. So, for example, e_{t-1} and e_{t-1}, etcetera, and in such situations, the covariance is reduced to the corresponding variances, OK.

So, here the overall covariance of this structure becomes θ 1 and then σ square e, and if you notice, all the other product or all the other cross-product terms would be 0s, right? Now, similarly, can you extend this to one more lag?

So, let us say γ_2. So, γ_2 means what? So, γ_2 is nothing but the covariance between yt and y(t-2).

$$\gamma_2 = Cov(Y_t, Y_{t-2}) = Cov[(c + e_t + \theta_1 e_{t-1})(c + e_{t-2} + \theta_1 e_{t-3})] = 0$$

So, the lag should be 2, right? Now, again, I can do the same thing. So, I can replace yt here and I can replace y(t-2) here and then talk about the covariance of yt, y(t-2). Now, something interesting one can notice here is that you cannot find the product of error terms where the subscripts are the same anywhere. So, essentially even if you take any product, for example, e_{t-1}, e_{t-2} or e_{t-1} from here and then e_{t-3} from here or e t and then e_{t-2} from here.

So, all such combinations, the subscripts are different. So, covariances of error terms where the subscripts are not the same happen to be 0. So, essentially the covariance here happens to be 0. So, thus this is a very important property. So, γ_k happens to be 0 for all lags or for all k strictly bigger than 1.

$$\gamma_k = 0 \text{ for all } k > 1$$

And when k equals 1, we have seen here that γ_k is nothing but, how much? So, γ_k is nothing but θ 1 into σ square e. So, when k is 1, this is the answer for the autocovariance and for every other lag above 1, the autocovariance happens to be 0. So, using these two facts, I can actually derive the ACF. So, ACF kind of takes a very simple form.

So, whenever k is 1, so by the way ACF is ρ k and then ρ k is nothing but γ_k by γ 0. So, whenever the lag is 1 right. So, if you remember how much was γ_k. So, γ_k was nothing but θ 1 into σ square e, right. And if you remember γ 0. So, γ 0 is nothing but the variance function of a MA 1 process.

So, if you go back a couple of slides. So, this is exactly the variance function of a MA1 process. So, σ square e into 1 plus θ 1 square which is nothing but γ 0, right. So, if we replace these two quantities in this ratio here. you will essentially get what you see here.

So, γ 0 is how much? So, γ 0 is σ square e into 1 plus θ 1 square. So, if you take the ratio of γ_k and γ 0, you get what you see here. So, whenever k is 1, the ACF function is θ 1 divided by 1 plus θ 1 square, and for that matter, for any other lag above 1, the ACF has to be 0 because the corresponding γ_k is 0. So, the ACF function of an MA 1 process is very easy.

So, it takes some value, which is this value at lag 1, but for all other lags above 1, the value is strictly equal to 0. So, now, can you extend this to MA2, and then we will see what happens. So, for example, the MA2 process takes that form, right? So, we extend one more lag. So, yt equals C plus e t plus θ 1 e_{t-1} plus θ 2 e_{t-2}, ok.

Now, the mean function happens to be c again, as we saw earlier, while the variance function happens to be σ square y, which is γ 0, which is nothing but σ square e into 1 plus θ 1 square plus θ 2 square. So, I can actually blindly apply the formulas which were derived earlier on the very first slide, right? So, the overall mean again becomes c, and the variance function happens to be that thing. So, now, how about the ACF function? Right. So, for that, I can again construct the different auto covariances one by one and then eventually see if we get some pattern or not.

So, again, γ_1. So, γ_1 is what? So, γ_1 is auto covariance at lag 1. So, covariance between yt and Y_{t-1}, ok. So, again, this entire thing is yt, and then this entire thing is Y_{t-1}.

$$\gamma_1 = Cov(Y_t, Y_{t-1})$$

$$= Cov[(c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2})(c + e_{t-1} + \theta_1 e_{t-2} + \theta_2 e_{t-3})]$$

$$= \theta_1 (1 + \theta_2) \sigma^2_e$$

So, we want to find out the covariance between yt,Y_{t-1}, ok. Now, again, you apply the exact same logic. So, wherever you see the same subscripts or similar subscripts, you combine those terms, and then the corresponding covariance reduces to the variance and for all other combinations, right. So, for all other combinations, the covariance reduces to 0. So, here, where exactly can you see common subscripts?

So, for example, e_{t-1} e_{t-1}, right, and then e_{t-2} from here and then e_{t-2} from here. So, only two terms. So, θ 1 e_{t-1}, e_{t-1} reduces to something like θ 1 into variance, which is σ square e. And then θ 2 e_{t-2} and then θ 1 e_{t-2} reduces to θ 1 into θ 2 and variance, which is σ square e. So, probably, if you are finding it difficult, I can write it down. So, there will be one term which is the covariance of θ 1 e_{t-1}, e_{t-1}, right, and then plus the covariance of the other term, which one? So, θ 2 e_{t-2}, θ 1 e_{t-2}, ok.

So, this covariance is nothing but θ 1 into the variance of e_{t-1}, which is σ square e, and then this term is nothing but you multiply the coefficient. So, θ 1 θ 2 and then the variance of e_{t-2}, which is again, as per the assumption, σ square e, ok. So, if you add these two terms, right. So, if you add these two terms, this is exactly what you will get. Now, extending this to one further lag.

So, let us say γ_2. So, what exactly is γ_2? So, γ_2 is nothing but the covariance between yt and then y(t-2). So, this entire thing is yt, and then this entire thing is y(t-2). So, again, the same logic.

$$\gamma_2 = Cov(Y_t, Y_{t-2}) = Cov[(c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2})(c + e_{t-2} + \theta_1 e_{t-3} + \theta_2 e_{t-4})] = \theta_2 \sigma^2_e$$

So, wherever you see the same subscripts, they contribute to the variance; if the subscripts are different, they contribute to 0. So, for example, here, can you spot the same subscript? So, let us say θ 2 e_{t-2} and then e_{t-2}. So, at the end of the day, we will have covariance between θ 2 e_{t-2} and e_{t-2}. So, this would be nothing but θ 2 multiplied by the variance of e_{t-2}, which is σ square e, which is exactly what you see here.

And similarly, if you follow the same pattern and then keep on extending the lags, let us say γ 3, γ 4 up to γ_k, you will soon notice that none of the subscripts are common, right? So, the subscripts in the first term and the subscripts in the second term would be entirely different. So, what do you mean by that? So, all the further γ_k's after 2 would be 0, OK? So, this is a very important statement.

So, just to combine whatever we have learned in this session, the auto-covariances and eventually the ACF function for a general MAQ process vanish for k, which is strictly bigger than q, right? So, if k happens to be, let us say, q plus 1 or q plus 2. So, at that particular lag, all these functions would be 0. So, for example, γ_k is let us say γ q plus 1 or γ q plus 2, and so on. So, at all lags above q for an MAQ process, the ACF functions and the auto-covariances would be exactly equal to 0.

So, again, here we have some illustrations. So, some simulations, kind of a thing. So, let us say in the first plot, what we have is a plot of ACF for a particular MA1 model where again the coefficient happens to be 0.6. So, 0.6 is nothing but the value of θ 1 here, OK? Now, notice one thing.

So, this plot is nothing but the correlogram or the ACF plot. So, if you go back a couple of slides, we have the function of ACF for MA1, right. So, we derived that. So, this is exactly the ACF function for the MA1 process right here, and the value is what? So, if you again remember, the value is θ 1 divided by 1 plus θ 1 square.

So, now, again, if you go back to that graph, we will see where do you see the first spike. So, obviously, at lag 0, the value would be 1 always, right, because the ACF function at lag 0. So, ρ 0 happens to be 1 because the correlation between yt and yt itself, because the lag is 0. It is always 1, right? And for all the further lags, for example, 1.

So, whenever the lag is 1 here, so the value was what, you remember? The value was θ 1 divided by 1 plus θ 1 square. And then, this should be my ρ 1. Now, in this experiment, θ 1 is 0.6. So, this value should be nothing but 0.6 divided by 1 plus 0.36, which is nothing but 0.6 divided by 1.36, which is kind of close to something like 0.44 or something like that.

And then here, clearly, we see that at lag 1, the value is 0.44 somewhere here. And similarly, if you see for all the future lags, so 2, 3, 4, 5, the value happens to be 0. So, we derived that. So, whatever we derived is kind of reflected in this ACF plot of an MA1 process. So, now, probably in the next lecture, what we will see quickly is how to identify particular models based on several plots.

Particularly the ACF plot and the PACF plot. So, if someone is trying to analyze a time series dataset and then practically work on that. So, I think this is a very, very important exercise based on certain ACF plots and PACF plots. Can you kind of identify the underlying model or have a very strong guess as to what the underlying model could be? And one can actually do this for some very basic models.

So, for example, here you see AR2 or MA2, right? And then the top row gives you ACF plots; the bottom row gives you the PACF plots. So, just based on these four plots or based on a collection of ACF plots and PACF plots and by observing the patterns there, can you kind of identify that a possible model could be one? So, again, like I said, the ACF plots and PACF plots are really important tools in model identification. So, probably in the next lecture, we will kind of continue with this and then we will bring in the idea of a random walk.

So, a random walk is a very easy idea. So, probably we will discuss some simulated random walks and then we will derive some very basic properties of a random walk. Thank you.