Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 36
Lecture 16A - Autocorrelation and Partial-autocorrelation Functions with R Demonstration

If you recall what we have learnt is the concept of auto correlation function and partial auto correlation functions and I am sure now the concepts are a lot more reinforced now that you have written your quiz and hopefully you know all the requisite concepts not only of the ACF and PACF, but the surrounding ones such as stationarity and so on, are better understood of course, one keeps developing prospective on these concepts, as one keeps going along even now 10 to 12 years into teaching this course, I still keep gaining fresh prospective, but that is me, but in general it is true that you keep developing prospective, but the basic understanding should be in place.

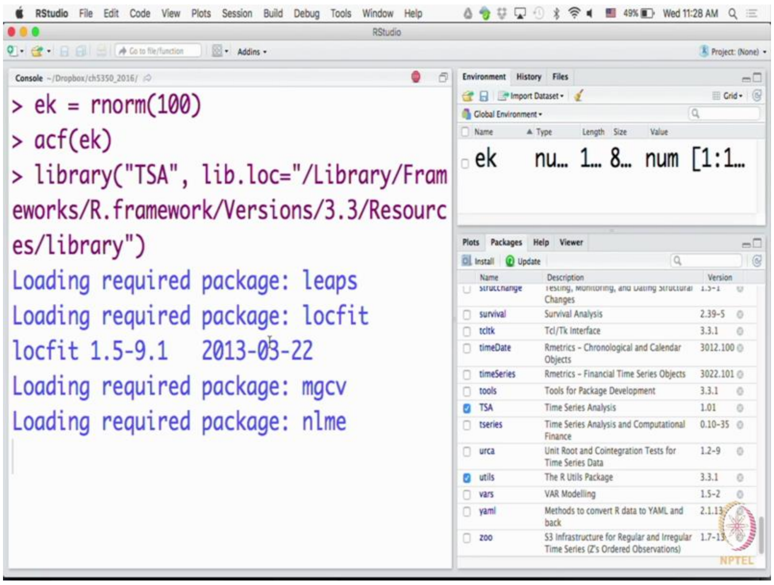
Now, if you recall we have introduced the auto correlation function and the PACF mainly to see if there is going to be is a scope for predictability within a series, that is the first objective and if there is predictability, we look at or the we examined a signatures of ACF and PACF and kind or reverse guess the model that must have generated the series. We have not talked about you know the mapping between the ACF and the models in a rigorous way, in other words we have not asked if there is a one on one mapping; if I see a certain signature in ACF, does it necessarily mean that only one model can generate it or there are multiple models possible for a given signature, all of these becomes necessary because we are going to work backwards.

So, we have gone through part one where we have learnt a measures, now part two is to understand models themselves and in that process we will also study what are the their ACFs signatures of this models, we are not going to different discuss completely different models from what we have seen already. We have already seen the class of moving average and auto regressive models, we are going to probably go a step further and discuss ARMA and ARIMA, and finally we will have the SARIMA. So, that is where will be heading in. In the process of studying this models, we will do two things we will study the theoretical properties of this models, in the sense under what conditions do this models generate stationery processes because that is of interest at as of now.

When we go to ARIMA we will incorporate processes that have random walk type non stationarities and then we move to SARIMA, we will also include processes which have seasonal components, but we begin with stationery processes and then we study the conditions the theoretical conditions on this models, that will allow us to generate a stationery process because there is no point in just considering a model without worrying about this finer aspects.

Now, before we do that I also want to give you a feel of what you would see in practice, in the sense if I am given a series, what is that you would actually do? Step one is of course constructing ACF right, now we have studied the theoretical ACF, but in reality the ACF that or the PACF that you see would be quite different from what you we have studied theoretically.

(Refer Slide Time: 03:33)

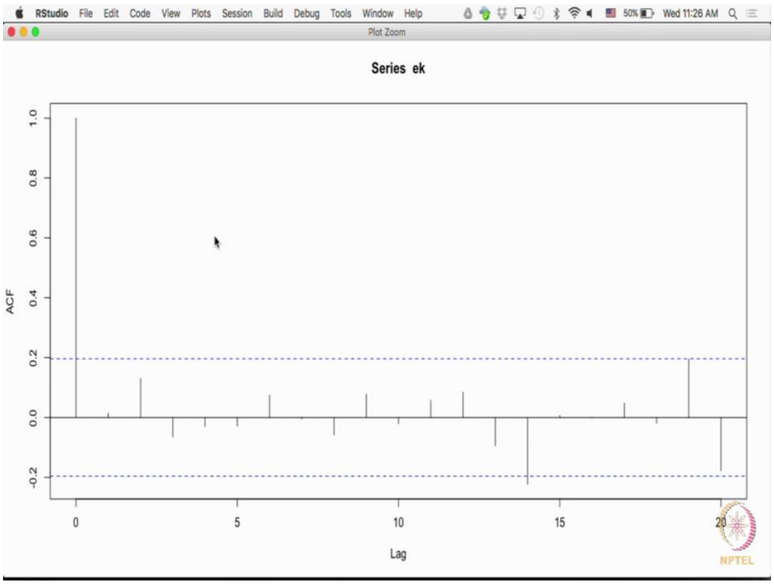


So, as an example suppose I pick some series that let us say white noise itself. So, I generate a white noise series of let us say 1000 observations or even 100 observations let us say; now when I use r norm I am generating a Gaussian white noise sequence in terms of r the way r looks at it is not as a time series, in other words this e k that I have generated is not the time series object at.

It is just a bunch of numbers that have been sampled randomly from a Gaussian distribution; the fact that I said randomly random sampling means that these bunch of numbers that we have generated are un correlated always, whenever you see random sampling it means that no two observations are correlated so; obviously, I can think of this as a white noise sequence, if you want to convert this to a time series object then you can coerce or force e k to be time series object, we will worry about that later on; as of now the command ACF in r, which estimates the ACF using the estimation expression can be applied to time series objects as well as to a vector.

So, this is a vector e k is a vector of random numbers sampled from a Gaussian distribution, now before we plot the ACF, we should recall the theoretical ACF that we have seen for a white noise processes, which is impulse right and when we plot it we should ask if really it matches with what we have learnt theoretically when we plot the sample ACF.

(Refer Slide Time: 05:14)

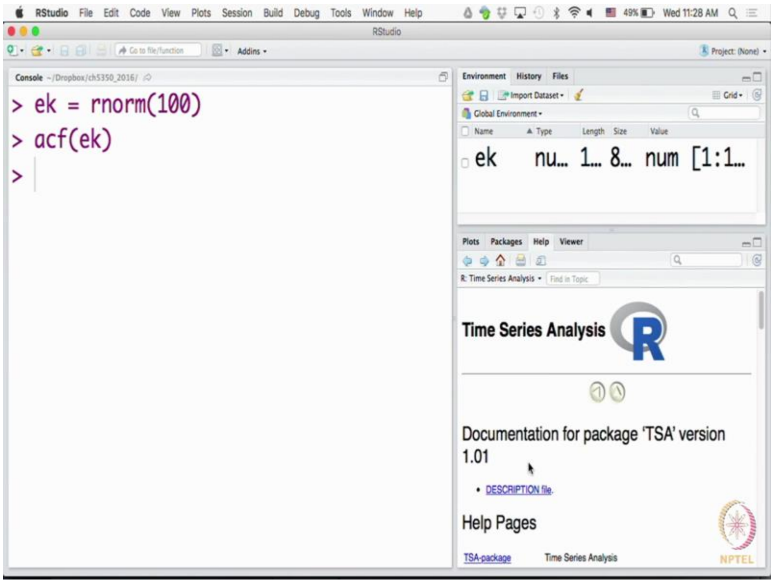


Now, this is how the plot looks like let me actually zoom out and show you the plot here. So, this is how the ACF of a sample realization or a realization of a white noise process looks like, where we have used 100 observations. Now you should learn to look at the ACF plot carefully particularly the x axis, the la the x axis begins from lag 0 there are different packages in r which plot the x axis slightly differently for example, there is a package called a s t s a let me show that you here I do not know how well you can see, but there is this package called a s t s a sitting at the top here and this is the full form of the package the name is applied statistical time series analysis, this is provided by the authors of the book applied time series analysis that is one of the references in your course by shumway and stoffer right and you can download this either from their website or it is available in your r distribution, simply when you install ask for a s t s a and it will install all the libraries, it is a good package to have and one or two assignments will refer to a few datasets from this package.

So, I recommend that you install this package and if you go into the package there are a number of routines here including some rewritten ACF codes are wrapped around one, I do not think people sitting at the far can see this clearly, but go back and on your desktop go and look at the list of packages available in this routine.

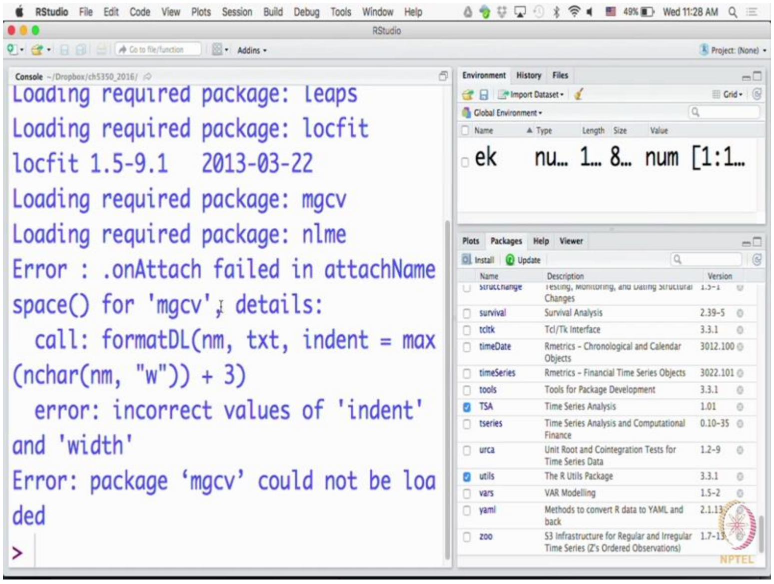
In this package for example, if you were to look at the ACF, I do not know how well you can actually see from the back; there are the let me go back to another package not a s t s a, but there is another package that you should install which is called the TSA; TSA stands for time series analysis.

(Refer Slide Time: 07:24)



If you click on TSA that is the package name after you have installed, you will see that there is another routine called ACF it bears exactly the same name and this is going to be the case in our engine wheel when you install packages, there are going to be some conflict of names and when you up when you sorry load a particular package for example, sorry when you load TSA.

(Refer Slide Time: 07:56)



For example it would say that there is some problem with the some package because my packages are not up to date, but if it loads properly then it should not it will tell you what are the routines that have been masked in the sense of the previous packages and one of them is ACF.

So, anyway going back to the help on ACF here let me go here. So, this is the ACF coming from TSA, I do not know how many of you can see at the top, but there is a this name is ACF and then bracket TSA. If you look up the help it will tell you essentially that it is a same one that we have just used from the base package, but with some slight difference in terms of masking the lag masking the ACF at lag 0. So, for example, here in the plot here if you see the ACF is plotted at lag 0, some authors believe and some analysts believe that there is no point in plotting ACF at lag 0 it is anyway known to be 1. So, why not plot from lag one onwards. So, you should be careful make sure you read the first lag at which this ACF is plotted.

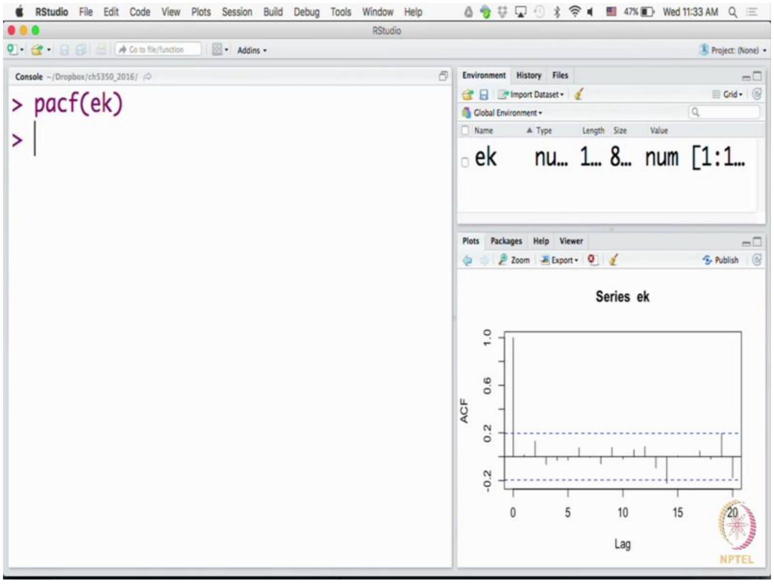
So, coming back to the discussion, the ACF should ideally have looked like an impulse but that is for the theoretical one it looks at the full ensemble, it essentially the entire the population, but here we have estimated ACF from just a sample of 100 observations; obviously, the this estimate is going to be different from the theoretical one, but it cannot be significantly different; now what is significant cannot be subjective cannot be an individual's opinion, one has to use statistics to figure out what is meant by significant significantly different dense on? As an example here ideally we should have expected the ACF at non 0 lags to be 0, which it is not right and also you can see these dash lines are essentially going to help us will learn about learn on how to construct this das lines?

What is the expression is there is there a statistical procedure or an expression to draw this dash lines and what do they mean? At the moment you should understand that they are called significance levels and a very crude interpretation at the moment I can give you is that, if these estimates that you are looking at fall within those bands that significance band then for all practical purposes you treat them to be insignificant. So, any estimate that falls within that band is to be treated as statistically insignificant, but that is a crude interpretation a much better and a clearer interpretation I will give you later on when we talk of in estimation when we discuss estimation theory.

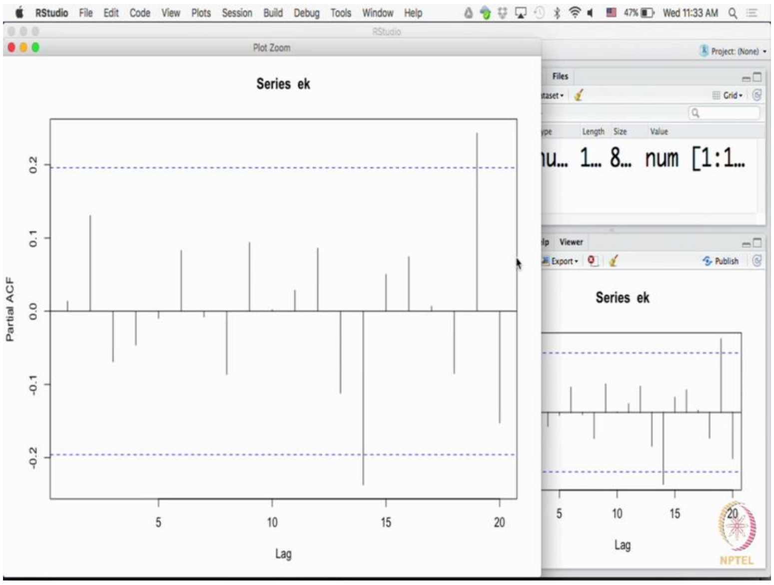
So, the point that I am trying to make you here is, soon you will be examining the ACF of certain data sets and you should now learn to draw a difference between the theoretical one and the sample one. The theoretical ones are only going to guide us, but what is important for us in practice is the estimates and of course, this is one way of estimating ACF, there are many other ways of estimating ACF, but this is a very commonly used estimator and the same discussion applies to PACF as well.

So, in other words theory will give you how the signature should look like, but in practice estimates can looks quite different not completely different, but quite different, but these kinds of significance levels will allow us to test hypothesis for example, the hypothesis that I would like to test for a series is that the given series falls out of a white noise process to begin with. How do I test that hypothesis? One way we will learn other ways of testing for whiteness, one way is to draw the significance bands and see if the ACF that is the sample ACF at all lags fall within that significance lag, that is they are statistically insignificant; there are more there are better ways of testing for whiteness, but this is a quick and easy way of testing for whiteness although it is not very rigorous

(Refer Slide Time: 12:43)



(Refer Slide Time: 12:49)



And the same applies to PACF as well. If I were to plot the PACF you would see the same thing, but notice now let me zoom out here, notice now that the PACF is actually not the PACF at lag 0 is not being shown. So, it is a strange thing if you were to plot ACF and PACF one below the other, then you would have ACF drawn at lag 0, PACF not being drawn at lag 0 and so on. So, to fix this kind of visual discrepancies some authors rewrite this ACF and PACF so that is the same lags are displayed and one of them is a s t s a which tries to do a wrapper I think there is a routine either it is a part of a s t s a or you can download it from Shumways Stoffers website, there is a routine called ACF 2 just to help you distinguish between ACF which drops at lag 0.

On the other hand in TSA, there is an option for the ACF routine in TSA package, where the user can say do not show me the ACF at lag 0. There is one point that I have not mentioned with regards to PACF; although PACF at lag 0 is not defined, you can you define it to be 1, just to be consistent with ACF right, but theoretically you will not find any PACF definition at lag 0 because there is no need anyways that is the point. So, this PACF kind of suggest that it is white noise, but you can see at the lag 19 there is an estimate that is going out of the significance (Refer Time: 14:17). So, now, the question is whether this is white, but then this can change with the realization also right and this is only one of the ways of testing for whiteness, there are better ways of testing as I said. So, this may give you some misleading results in any hypothesis test as you must have seen in the lectures there are going to be type one errors, there are going to be type two errors this the and there is a power of a test.

So, this way of testing for whiteness is not as powerful as other methods of hypothesis testing. One of the other ways is to fit a model and see for example, now we have seen that there is some predictability based on PACF, maybe I am compelled to fit a model, when I fit a model I should look at the errors in the parameter estimates and make sure that all the errors are significant anyway so much for the sample ACF and PACF.