Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 09

Lecture 43: Spectral Density Estimation

Hello all, welcome to this course on time series modeling and forecasting using R. Now, again, if you vaguely remember, if you are following this week, we are talking about the spectral density function and Fourier transformation. So, the focus idea for this week has been transitioning from a time-based approach to a frequency-based approach. Now, just in the last lecture, if you remember, we studied entirely about the spectral density function, or in short SDF, some of its properties, and some implications of SDF. And towards the end of the last lecture, we talked about a few examples.

So, white noise, AR1, MAQ, and then random walk. And for each of those examples, we laid down the expression for the underlying SDF. So, now in today's lecture, before we start with anything new, I will start today's lecture by mentioning a few practical applications of spectral analysis in general. So, we will start with signal processing. So, the first one is signal processing.

So, spectral analysis is fundamental in signal processing for understanding and manipulating signals in the frequency domain. So, the underlying few examples are as follows. So, noise filtering. So, let us say the first application could be to identify and remove noise from a signal by filtering out high-frequency components using the HDF, right? Now, let me just take a pause here and try to explain the first example because there has to be some practical component also, right?

I mean, simply studying theory would not help. So, how do you connect theory to applications or to practice? That is the important part, okay? For example, again, if you reread this example: identify and remove noise from a signal by filtering out some of the high-frequency components using the SDF. So, again, if you remember a couple of pointers from the last section where we discussed that, I can actually chart out the SDF, isn't it? Because I can chart it out or graph out the path of the underlying spectral density function.

And if you remember vaguely, whenever you see peaks there—whenever you see peaks in the underlying SDF—these peaks correspond to significant frequencies. So, the underlying significant frequencies are these and these. So, hence, once you chart out the SDF, once you create the SDF, once you graph out the SDF, wherever you see peaks in the underlying SDF, you can immediately say which frequencies are significant. And once you identify such frequencies, one can actually filter them out or identify and remove noise from a signal by filtering out some of the high-frequency components using the HDF. Because what might happen is, if you are transmitting a signal, all these peaks or all these high frequencies—or wherever you have peaks in the HDF—could mean noise.

So, think of a situation where you are listening to some radio, and while listening to some song on the radio or some news on the radio, suddenly, you might have witnessed that there is some disturbance, right? And then, whenever you see you have disturbance in that transmission, there is some extra noise there, right? So, you have some extra hush there or, you know, some extra blast, whatever, right? So, all these are noise components, and one might want to remove all the noise components.

So, whenever you see increased activity in frequencies, you want to filter them out. Hopefully, the idea is clear. So, here the spectral analysis comes to the rescue. So, this is exactly my second example. So, cleaning audio recordings or improving image clarity.

So again, if you have an audio recording and suddenly see a spike there, that spike could mean some noise. So, noise means some extra volume suddenly, something like that. So, all such noise components are unwanted, and you want to filter them out. So, both these examples come under the heading of noise filtering. So, the first application is signal processing.

There could be another application in telecommunication. So, the idea about modulation and demodulation. So, the first example is that spectral techniques analyze and design communication signals. Let us say AM, FM, or digital modulation, etc. And what could be one of the examples there?

So, ensuring that transmitted signals occupy specific frequency bands only. So, you are sort of restricting all the transmitted signals to particular or a few bands there. Okay.

Then, the third example could be entirely different, coming from the finance and economics area. So, let us say spectral analysis could be used to detect periodic patterns, trends, and cycles in financial and economic time series.

A couple of other examples within that area are market cycle detection. So, let us say identifying cycles in stock prices, interest rates, or GDP growth rates, etc. And what could be one of the examples there? So, using spectral density to uncover business cycles or seasonal patterns. So, this is a particular example which is not coming from signal processing.

So, a slightly different area which is again of very much relevance. And probably one last area could be, let's say, medicine and biology. So, how does spectral analysis come into play there? So, spectral analysis is crucial in analyzing some biomedical signals. For example, ECG or electrocardiography signals or ECG signals.

Which are sort of helpful to detect abnormalities in heart rhythms, right? By identifying irregular frequencies. So, the heartbeats could be irregular, right? And this is a problem, right? So, the example could be diagnosing arrhythmia.

So, arrhythmia means irregular heartbeats, right? So, a heartbeat is nothing but the heart transmitting some signals, right? So, can't you propose a model based on the frequency domain, identifying the frequencies and then filtering out unwanted frequencies to make it smooth? So, diagnosing arrhythmia is one application in medicine and biology. So, hopefully now you have the motivation as to why one must transition from time-based modeling to frequency-based modeling.

So, now the key idea in today's lecture will be spectral density estimation. So, once you formulate the SDF, once you formulate the underlying model containing all the unknown parameters, the remaining idea is how to estimate the underlying parameters in the SDF. So, here we will broadly discuss spectral density estimation in general. So, here we will break it down into two parts. So, the first part is parametric estimation.

So, parametric estimation means wherever you have some unknown parameters in play and probably you might know the underlying distribution as well. So, in general, parametric estimation in statistics means that you can. So, if you are able to put forward some statistical distribution, then it is called parametric estimation. And on the other hand, if you are not able to put forward or conclude what distribution it should follow, then we have an idea called non-parametric estimation. So, here we will clearly break it down into two parts.

So, parametric estimation and then non-parametric estimation. The first one is parametric. So, parametric methods assume a specific model for the time series. So, as discussed, a short while back, if you happen to know the underlying model or if you can sort of put forward a specific model for the time series, one can actually use the parametric umbrella of estimation. So, what do you mean by modeling?

So, the models could be as basic as, let us say, ARMA, ARIMA, etc. Again, parametric methods assume a specific model for the time series, such as ARMA, ARIMA, etc., and estimate the parameters of that model to compute the spectral density. So, if you are able to put forward, let us say, either ARMA or ARIMA as a model on the time series, then we want to estimate the underlying parameters of the ARMA model or the underlying parameters of the ARIMA process, etcetera. And again, probably many of you might know that, or should know rather, what the parameters in the ARMA model are. So, of course, the phi coefficients and the theta coefficients and probably the sigma square, which is the variance of the error component.

Similarly, ARMA what are the parameters? Same. So, all the phi coefficients, all the theta coefficients and probably sigma square which is variance of that E t or epsilon t basically ok. So, what are the steps here? So, the first step is model selection.

So, of course, we have to select or choose a particular or appropriate model. So, choose a model class, let us say ARP or MAQ or let us say ARMA PQ based on the characteristics of the time series. So, the first idea, the first point under the steps is model selection. So, choose a model class such as ARP, MAQ or ARMA PQ based on the characteristics of the time series. Second is parameter estimation.

So, estimate both the parameters let us say phi and theta of the chosen model using usual methods like MLE etcetera. So, wherever possible one can apply MLE or one can actually apply method of moments also right. So, whichever applicable methods are there for parameter estimation of that particular model one can actually apply. So, a more general looking method is rather MAD, right. So, let us say so, the first step is model selection.

So, let us say you choose an ARMA PQ model, and now the second step is parameter estimation. So, you want to estimate all the underlying coefficients. So, all the phi coefficients, all the theta coefficients, using appropriate techniques—let us say MLE, okay. Then the last component, or the last step, is to compute the spectral density function. So, once you formulate a model and estimate it completely, then you can actually write down the SDF.

So, use the analytical formula to compute the SDF of the chosen model. Make sense? So, what exactly are the pros? So, some of the pros are that it provides smooth estimates, right? So, parametric estimation.

So, the pros of parametric estimation are that it provides smooth estimates. The second advantage is that it is suitable for well-modeled stationary processes. So, wherever you know the actual model—let us say ARMA PQ, ARIMA PDQ, AR, MA, SARIMA, etc. So, whenever a particular complete model is defined, or rather well-defined, exactly there you can use a parametric estimation technique, right? Because if you know the model, if you can estimate it using MLE or the method of moments, then why should you go ahead with non-parametric, right?

So, parametric estimation is suitable for well-modeled stationary processes. But at the same time, there are some disadvantages as well, or what exactly are the cons? So, it requires the correct model specification, right? So, the second advantage here could also be a disadvantage, that it has to rely on some well-modeled stationary process. So, the correct model specification should be known.

If it is unknown, then of course, you cannot proceed with parametric estimation, or if you are not very clear about, let us say, what the model could be or what the model would be, then parametric estimation is not the solution. And the second con, or the second disadvantage, is that it may not work very well for complex or unknown processes. So, if the overall time series behavior is complex or the overall time series behavior is unknown, right? Or you have a combination of different, let us say, two different time series, something like that. So, where the overall structure is slightly complex, parametric estimation might not be the solution, okay?

So hopefully, the first set of estimation is clear: parametric means that some underlying model is known to you, and you are able to estimate the parameters of that underlying model using MLE, method of moments, etc. And once you estimate the model, the last step is to create that HDF from the analytical formula. Now, on the other hand, we will talk briefly about non-parametric estimation. So, non-parametric methods do not assume, as discussed earlier, non-parametric approaches do not assume a specific model for the time series and directly estimate the HDF or spectral density from the data itself.

So, here we are not saying that the underlying model is ARMA or ARIMA—nothing like that. And hence, we have a non-parametric approach. Now, we have a very famous approach underlying the non-parametric framework, which is called the periodogram. And, of course, in almost any famous textbook on time series or spectral analysis, you will find the idea of a periodogram. So, what exactly is a periodogram?

So, a periodogram is nothing but a fundamental non-parametric method for estimating the HDF or the spectral density function. So, one can actually estimate the HDF through a periodogram. And a periodogram provides a measure of the power. So, power is nothing but variance. So, the power of a signal.

When you say the power of a signal, the power of a signal is nothing but the variance of a signal. So, a periodogram provides a measure of the power or variance of a signal at different frequencies. Now again, I can graph a periodogram. So, a periodogram is nothing but a graph, by the way, with sort of involving the significant frequencies and what is the variance of the underlying signal at those significant frequencies.

So, do not worry. We will see some examples down the line where we will see some periodograms, and obviously, in the R session towards the end, we will bring in and tie all these details and then sort of put forward a common practical approach there. But again, like I said, I can actually graph or chart out a periodogram. So, a periodogram can be graphed out, and in that graph, one can actually identify the significant frequencies of the signal having some different powers. Okay? But then, entering into small technicalities, I can actually define a periodogram. So, how do you define a periodogram?

So, let us say for any discrete time series Yt of length capital T, Yt is assumed to be a discrete time series of length capital T. The periodogram at frequency omega is defined as follows. And exactly how? So, I(omega) is nothing but 1 by capital T absolute value of summation T going from 1 to capital T Yt into e to the power minus I omega T and then whole square. So, in a way, all the formulas are kind of slightly similar, right? So, even from the earlier lecture, the last lecture, we defined S(omega), right?

$$I(\omega) = \frac{1}{T} \left| \sum_{t=1}^{T} y_t e^{-i\omega t} \right|^2$$

$$I(\omega) = \frac{1}{T}[Re^2(\omega) + Im^2(\omega)]$$

So, even there, the S(omega) formula involves some summation, some integrals, and then this component was there. So, e to the power minus i omega h, and then here we have t. So, you have slight differences between the formulas for S(omega) and I(omega). Now, again here, the whole difference is it is not 1 by 2 pi, but rather 1 by capital T. So, capital T is the length of the time series, right? And then again, you have some absolute value summation running from 1 to capital T, unlike minus infinity to infinity there, right? So, all these are some subtle differences, but my point is that this component that you see here almost appears in all the formulas.

$$\omega_k = \frac{2\pi k}{T}, \;\; k = 0,1,\ldots,T-1$$

So, e to the power minus i omega into something, right. Anyway, so this is the formula for getting the periodogram, which is denoted by capital I of omega, where again, hence I am mentioning that all the notations are as before. So, omega is a frequency, small t is nothing but a time point, capital T is a length, and y t is the underlying time series. Alternatively, I can write down the formula for the periodogram slightly differently. So, I omega could be written down in this manner as well, and how so? 1 by capital T into real part square plus imaginary part square.

So, what I am doing here is that since again this formula involves this exponent here or exponential term, and as seen before, you have a relation between this exponential term and sine function and the cosine function, okay. So, cannot I split that into, let us say, the real part and the imaginary part? The answer is yes, okay. So, I of omega is 1 by t square of the real part plus square of the imaginary part, where the square of the real part is nothing but summation yt into cos of omega t, and square of the imaginary part is summation yt into sin of omega t. By the way, all these things could be easily proved if somebody is really good in mathematics or, in general, let us say, real analysis or complex analysis rather. So, he or she should know the relationship between this guy and then sine and cosine in some sense.

So, in general, you have these two kinds of alternative formulas for defining the periodogram at a particular frequency omega. Now, again, one important point to note here is that beat S omega beat I omega. So, again, just for a second, let me go back to SDF. So, S omega denotes the spectral density function. So, this is nothing but a function in terms of frequency.

Can you see that? Because omega is a frequency. So, the spectral density function is nothing but a function of the frequencies. So, again, if you graph out the SDF or if you graph out the periodogram, on the x-axis you will have the frequency. So, be it a graph of S omega or be it a graph of I omega.

Make sense? Because both are functions of omega in a way. I think this important but small point I missed in the last lecture is that both SDA, the spectral density function as well as the periodogram, are nothing but functions of the underlying frequency. Make sense? So, now what exactly are some properties of the underlying periodogram?

So, the first property is the frequency range. So, which means that the periodogram is evaluated at discrete frequencies. So, let us say omega k, and then omega k equals 2 pi into k divided by capital T, and of course k can range from 0, 1, 2, 3, 4 up to T minus 1. So, of course k cannot be T because T is the last threshold, because T is the number of Because small t goes from 0 to capital T, right?

So, k can't be t and k can't go beyond t. So, you have a limitation with k. So, it can run from 0, 1, 2, 3, 4 up to t minus 1. But the periodogram is evaluated at discrete frequency. So, for each and every frequency, if you replace the values of k one by one here, I can actually evaluate the periodogram at those frequencies, which are omega k. And second point is for real value time series, the periodogram is symmetric, ok. So, for a real value time series, periodogram is symmetric.

So, I can actually restrict my frequency range to only 0 to pi rather than 0 to 2 pi. Because if you have any real value time series, which in general we have, right, because all the time point or all the values in yt would be real valued. You will not have any complex observation. So, let us say temperature data, rainfall data, stock price data. So, all these values are real value.

So, in that sense for any real value time series the periodogram is symmetric. So, you can actually restrict the frequency range between 0 to pi only. So, these are some underlying properties of I omega or the periodogram. Then the units. So, the second idea is units again under the properties heading.

So, what do you mean by units? So, the periodogram is measured in units of variance per unit frequency. Again, just repeat. So, the underlying units for how one measures the periodogram are nothing but variance per unit frequency. For example, if the underlying time series is in volts, then the periodogram is measured in volts squared per hertz.

So, Hz stands for hertz. So, volts squared per hertz. And then the next one is bias and variance. So, what do you mean by that? So, the first idea is bias.

So, the periodogram is asymptotically unbiased. This is again one very important property. So, firstly, what do you mean by asymptotically? Asymptotically means if something converges to infinity or, in the long run, what happens. So, the periodogram is asymptotically unbiased, which means that the expectation of $I(\omega)$ is close to $S(\omega)$. So, the expectation of $I(\omega)$ is close to $S(\omega)$.

Again, if you sort of forgot that what we are trying to study here is the estimation of S omega. So, we began this session by—so again, if you go back a few slides, you will see that we had a heading here which said 'spectral density estimation.' So, I omega is nothing but an estimate of S omega. So, we are trying to estimate that spectral density using a periodogram. So, in short, the expected value of the periodogram is close to S omega.

So, here I am not writing exactly equal to because it is asymptotically unbiased and not exactly unbiased. Make sense, hopefully? So, the expectation of the periodogram should be somewhat close to my SDF. And lastly, variance. So, the periodogram is not consistent.

This is one drawback—probably then, the periodogram is not consistent. So, what do you mean by consistency? Consistency means that the variance does not decrease with increasing capital T. So, even if you collect more and more observations in the time series. So, if you go on increasing my capital T, still in that case, the overall variance does not converge to 0, or the variance does not decrease, okay? Which is probably a disadvantage because we would want the variance to converge to 0, right? If you collect more observations, then naturally the variance should go to 0, which is not happening here.

So, the periodogram is not consistent, but at the same time, it is asymptotically unbiased. So now we will pay attention to what exactly the steps are to compute the periodogram. So now that we have some idea about what you mean by periodogram. So, the periodogram is nothing but an estimate of the SDF. But how do you compute it?

So, the first step is to transform to the frequency domain. And here we will make use of the DFT, or the discrete Fourier transform, or discrete Fourier representation of the time series. Now, again, probably for this, you might have to revisit the first lecture of this week where we talked about Fourier transform, inverse Fourier transform, and all these things. But here, transforming to the frequency domain uses the DFT, or the discrete Fourier transform, of the underlying time series. So, y and then omega k happens to be equal to summation t going from 1 to capital T y t into e to the power minus i omega k t.

$$Y(\omega_k) = \sum_{t=1}^{T} Y_t e^{-i\omega_k t}$$

Then, the next step is one has to compute the power spectrum. So, there is something called a power spectrum. Now, what do you mean by that? So, it means that the periodogram is nothing but the squared magnitude of the DFT scaled by 1 by T. Again, the periodogram is nothing but the squared magnitude of the DFT scaled by 1 by T. Now, one small typo here is that This should be again Y of omega K, and then this Y of omega K is the same thing as this thing here.

$$I(\omega_k) = \frac{1}{T}[X(\omega_k)]^2$$

So, the first step is you find out the DFT of a time series. Then you somehow square the DFT, scale it down by capital T, and you can get the periodogram at that frequency. Hopefully, this is clear. Again, just to repeat, the first step is to find out the DFT of the time series using this structure here. And then, once you find out Y of omega K, you square it and then scale it by 1 by T, and then you get I of omega K. And then the third step is frequency resolution.

So, what do you mean by that? So, the resolution of the periodogram depends on the length capital T of the time series, and in turn, what do you mean by that? So, longer capital T or longer T provides better frequency resolution. And here we have a small relationship. So, triangle F—so this is a technical notation.

$$\Delta f = \frac{1}{T\Delta t}$$

So, triangle F equals 1 by capital T, triangle T. So, triangle F means a very small-time step or a very small interval. So, by the way, this notation is called a sampling interval. So, what happens in a very small-time span? So, triangle F equals 1 by capital T into triangle T. So, by the way, this step is kind of a technical step.

If you do not want to pay much attention here, you can sort of ignore this, but this is one step to compute the periodogram, which is called the frequency resolution. So, once you sort of formulate the DFT, then once you sort of connect the DFT to find out periodogram, last step is to sort of do the frequency resolution. Okay. So, once we understand that what do you mean by periodogram and then how do you get the periodogram. So, what are the limitations now?

So, some limitations of the periodogram. First limitation is a noisy estimate. So, the variance of I omega does not decrease with sample size t. We have seen this earlier. So, I omega is not consistent leading to noisy and unstable estimates. So, even the long run since the variance is not approaching 0 or variance is not reducing, we might observe some noisy estimates down the line.

Second is spectral leakage. So, what do you mean by that? So, when the time series contains frequencies not aligned with the discrete Fourier frequencies, power leaks into adjacent frequencies. There might be a case that since you are trying to estimate S omega using I omega, but the underlying S omega, the frequencies in S omega and the frequencies in I omega may not match exactly. So, in that case what happens is there is some spectral leakage.

So, there might be some leakage or power leaks into the adjacent frequencies. Hopefully this is clear that if both the graphs of S omega and I omega are not superimposing. So, let us say if you have this is the graph of S omega something like that and then let us say on top of that I am superimposing the graph of I omega and then here in I omega I can see that the peaks are slightly different. So there could be a situation where you have some spectral leakage. So the power leaks into the adjacent frequencies.

And lastly, resolution versus variance trade-off. So high frequency resolution comes at the cost of increased variance and vice versa. So if one needs some high frequency resolution, so let's say image clarity or signals or the pixels in an image, right? And if you want some high frequency resolution, that always comes at a cost of increased variance and vice versa.

So, if you want some high frequency resolution, if you want some immaculate resolution, then the variance also increases. So, you have a trade off. So, between resolution and the variance. So, hopefully in this lecture I have tried to cover a few ideas about periodogram and periodogram is one very useful, very important method of estimating the underlying spectral density. So, for that matter down the line if you have any practical data set that you want to sort of put forward a frequency domain based time series approach on that trying to put forward a model and during that procedure you would want to find out the SDF right because spectral density as seen in the last lecture is a key to this entire process of the frequency domain based approach.

Then, how do you estimate the SDF? So, the periodogram is one such estimator—a non-parametric one. You do not require any modeling here of the underlying SDF. Now, in the last theoretical lecture of this week, we will try to close the discussion about spectral density, periodogram, and spectral representation with some final points.

Thank you.