Spatial Statistics and Spatial Econometrics

Prof. Gaurav Arora
Department of Social Sciences and Humanities
Indraprastha Institute of Information Technology, Delhi

Lecture - 12C
Covariogram and Correlogram

All right. So, welcome back to this short part of lecture 12, where we are going to cover these spatial dependence devices called the Covariogram and the Correlogram.

So, let us dive in you know in all likelihood it will be a short lecture. So, consider the relation that the variance of Z of S_1 minus Z of S_2 is equal to the variance of ZS_1 plus the variance (I am simply applying the definition of variance here), of ZS_2 minus twice the covariance of ZS_1 comma ZS_2.

And remember you know these will be sort of true for all S_1 and S_2 in a stationary domain right? Now variance of ZS_1 minus ZS_2 is obviously, the variogram right. It is 2 gamma L_2 norm of S_1 minus S_2 right of course, you know we are talking about data that are delineated with a coordinate system where you know as we saw earlier S_1 is an x y coordinate and S_2 which is S_1 plus h is sort of also has its own x y coordinate and we can visualize these as vectors themselves right.

So, S1 and S2 are vectors and the lag between them is the vector h, which we say S_1 minus S_2. So, S h encapsulates both distance and direction right. So, we say that you know h the distance is equal to the L_2 norm of S_1 minus S_2. Now given S_1 S_2 and h as defined above Z has to be a second-order stationary process for the covariance you know covariance measure C of S1 minus S2 that is covariance Z of S_1 Z of S_2 to be defined.

Now, you know this C of you know the distance between S_1 and S_2 can also be written as C of h simply right? it is just C of h right?, and now I can write variance of Z of S_1 minus Z of S_2 is equal to C0. So, the variance of Z of S_1 is just the covariance of Z of S_1 with itself right plus C 0 minus 2 C h which can be written as twice C 0 minus C h right that is a convenient interpretation. Now, this variance of ZS_1 minus ZS_2 is equal to twice gamma h right?, where h is nothing but the distance metric of the lag vector between S_1 and S_2.

So, 2 gamma h is equal to 2 C 0 minus C h, which implies that gamma h is equal to C 0 minus C h right? quite clearly. So, Note 1, gamma, and C are inversely related, and this discussion cannot be clearer than that right? And Note 2 the fact that I have defined 2 gamma h means I am assuming that Z is an intrinsically stationary process. Remember to be able to define a variogram, I need the definition of intrinsic stationarity, right? The variogram is embedded in the definition of intrinsic stationarity.

So, I am just going to say that the fact that I have defined 2 gamma h means I have assumed or claimed or committed to intrinsic stationarity, which again remembers is a decision that you must take before, you go on to provide such a definition. Now Note 3, C of h this C of h is called the covariogram. So, we have solved one puzzle we know what is covariogram. Next, we will simply figure out what is correlogram right ok let us do that.

So, given C 0 greater than 0, we have rho, which is the notation for the correlogram for points separated by a lag of h. This is my correlogram is equal to C of h divided by C of 0, this is the definition of a correlogram. Now pay attention that the correlogram is an analog or analogous to the correlation parameter right?

So, if I were to do a little aside here and we want to talk about the correlation between x and y, those who have most of you would have seen this earlier is nothing but the covariance of x and y divided by the standard deviation of x and standard deviation of y right. We are simply writing an analog, we are defining an analog of the correlation statistic as we understand it. In fact, even define this as a row parameter using a row parameter.

So, we are simply sort of you know we are creating this analog of the correlation parameter. I am just going to fix my handwriting here sorry about this trouble. Yes, that is it alright; so, now sort of the properties portion. So, let us look at that. So, we have, if I am going to say furthermore if C of h approaches 0 as h approaches infinity.

So, we are looking at points that are very far apart or enough far apart in space right? then we say that 2 gamma h we will have 2 gamma h will approach 2 of C 0 right, which is nothing but you know basically the variance of ZS_1 and variance of ZS_2, the variance of ZS_i right? As obviously h approaches infinity.

Now, the quantity C 0 is called the SILL of the variogram. Notice that the SILL of the variogram measures large-scale variation in data right? It is a point when 2 gamma h has become a constant right? Remember when we looked at the properties of the variogram we said if you know 2 gamma h or the variogram becomes a constant k; that means, there is no spatial order correlation, spatial correlation, there is no spatial contiguity in data thus the data becomes spatially random.

Now, what this property with the covariogram tells me is that as we move far apart enough right? You are going to have a situation, where the data will be not related to each other, they will be random with respect to each other. So, spatial dependence by itself will sort of fall down, as we go as we increase h eventually sort of you know diminishing to 0. At this point the variogram measure will become a constant and that constant will be nothing but C 0, which is nothing but the large-scale variation in the data, the overall variation in the data.

Now you know a couple of more definitions that C of 0 minus C 0. Now this C 0 is the nugget effect and the big C 0, the capital C 0 is SILL. This is defined as the partial SILL. I am only defining these if you sort of come across these measures later and very importantly, so another very very important you know discussion is that the smallest value of a measure R. Now R is you know in on the real space. So, it is a one-dimensional measure just like h. So, it only encapsulates distance, not direction.

So, R which so the smallest R-value, for which 2 gamma R is a constant that is 2 C 0 right? then R is called or known as the range of the variogram or the semi-variogram. So, the range is the distance from any point in time where spatial dependence diminishes to 0 ok.

Spatial dependence is not a global process it is a local process. We have said that when we introduce spatial statistics we said there is spatial heterogeneity, which is large-scale trends in space and then there is spatial dependence, which is defined to be a local property right?

So, if we move far apart enough right? We are going to have a situation where these values are as if randomly generated from you know independently and randomly independently generated from each other. They are not going to be spatially dependent.

So, after we have moved out R units of distance from any given point the prediction power of this knowledge of the information of this original point beyond this R distance in all directions is 0, there is no value to this in terms of spatial prediction ok. This is pretty interesting, right? So, having understood that now what we have seen is that we have two alternative measures of spatial correlation right?

We have two alternative measures of spatial contiguity, right? We have the variogram we have the covariogram. We also have the correlogram, but the correlogram is really just you know a derivative of the covariogram. It turns out that the variogram is most favored among the two and the question is why? So, let us try and sort of you know understand what is gained by postulating a variogram rather than a covariogram.

So, typically you will see studies estimating or reporting variogram measures rather than covariogram measures whereas, we have seen on the previous slide, we saw that they are simply inversely related, they are linearly related in fact. But still, you know studies almost always will report a variogram rather than a covariogram and the question is why? The first reason is that the class of second-order stationary processes is almost completely contained in intrinsic stationary processes.

We know that for defining variograms we need intrinsic stationarity, but for defining covariograms we require second-order stationarity. Second-order stationery is a weaker concept of stationary it is completely contained in the intrinsic stationarity. So, that means, that the variogram is a more general device of spatial contiguity.

So, let us say that. The class of second-order stationarity is a process strictly contained within the class of intrinsically stationary processes. And that will basically imply that the variogram may be defined for more general settings than the covariogram is defined.

So, the variogram is defined for more general settings than the covariogram, an example of this is if you had to model Brownian motion in space. So, Brownian motion is nothing but again a collection of random variables in space basically forming a random function right? So, you have s in a domain D which is part of let us say a general D dimensional real space, then the variance of wS_1 minus wS_2 will be equal to just h, where obviously, h is nothing but S_1 minus S_2.

Now, this is a typical formulation of the variogram, now variogram is a function of the distance between two locations right? So, we are fine. So, instead of so now, gamma h is simply h right or h o basically 2 gamma h is simply h. So, gamma h is h over 2, right; whereas, the covariance of wS_1 and wS_2 is random variables at different locations S_1 and S_2. If you calculate it for the Brownian motion it will turn out to be half of S_1 plus S_2 minus S_1 minus S_2, for every S_1 S_2 in the domain D.

Now, S_1 minus S_2 is equal to h, but what we see here is that the covariogram also depends on location and not just distance. So, a covariogram is not defined in the sense of second-order stationarity. Second-order stationarity requires that the covariogram be defined by the distance between two locations and not the location itself. Here if I move on to locations S_3 and S_4 which are let us say apart by the same distance h, the covariogram will not be able to be the same as it is between wS_1 and wS_2 right?

So, the covariogram is not defined right. So, this implies the covariogram is not defined for a Brownian spatial process right? whereas, the variogram is because it is following the rule that you know the variogram the variance of the first difference of values that are generated at two different locations has to be a function of the distance and not the locations themselves.

It is not a function of S_1 and S_2, but S_1 minus S_2. The second major reason is something that we have not come to yet. So, we will just touch upon it and move forward that spatial prediction is considered easy to sort of carry out on intrinsically stationary domains, and intrinsic stationarity is strictly containing you know the idea of second-order stationarity.

So, that is why the variogram is more general we have seen an example. So, for prediction purposes, you know variogram is always going to be preferred over the covariogram that is why you know in most cases you know, you the papers that you will read, the applications that you will come across the applications that we will do in this class.

We will be working with the variograms and almost never working with the covariograms. We studied them for pedagogical reasons because we are going through the theory of spatial you know the estimation of spatial processes you know characterizing spatial processes are characterizing the spatial process. So, both variogram and covariogram provide a valid characterization of the spatial processes. The only trouble is that the covariogram is less general. So, they are not valid for cases or for situations, where the variogram maybe you know could be studied.

So, what we are going to do next is the next sort of agenda in this course is going to be working with what is called the experimental variograms. So, we are going to focus on the variogram of course, we have talked about it and we are going to sort of you know to focus on calculating or deriving this variogram using the real-world data set right? So, that is the agenda going forward, I hope you enjoyed today's lecture and I look forward to having you again in the next lecture.

Thank you very much for your attention.