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

Lecture - 14C
Variogram model fitting

All right, So, now, that we have spatial dependence in data.

The distribution of Z will look slightly different and we have seen that it will be different in the variance-covariance matrix representation. So, now, my sequence Z which is the sequence of data points which are n data points that I have, I am going to define it with a location vector s. So, instead of sort of you know working with the sequence Z_1 to Z_n, I am now sort of picking bringing in these location indices s_1 s_2 till sn just to sort of you know notational represents the data being a spatial data set right?

So, mathematical notation is just helping me give this extra rich dimension to my sequence right? I can say Z_1 or Zs_1, it will not really matter, but s_1 is the location it is deterministic and Z is the value realized which is a random draw from a distribution, right? So, it is a rich English interpretation of this mathematical device Zs_i. This is going to be distributed again as a multivariate normal with mean mu.

So, we have the same mean as the spatially independent case, but the variance-covariance matrix now is given as sigma with the functional dependence on theta; obviously, sigma is a n by n matrix. Now, this sigma theta provides me a measure of covariance of values at location Zs_i and location s_j right.

So, it provides me a general covariance structure between the data, and this structure you know is driven by their locational differences s_i and s_j. If I were to think about the matrix formulation well it's a n by n matrix. So, it's really a collection of all-you-know covariance Z_i Z_j is just an element where i and j we know go from 1 to n right. So, we are basically looking at different combinations of i and j in this covariance element right?

So, when s_i and s_j are represented in the same location that is I, is exactly equal to j, I have the diagonal element. So, I have a variance of Z_1, a variance of Z_2 all the way to a variance of $Z_{n,}$ right? The off-diagonal elements are interesting now, I have covariance Z_1 and Z_2 covariance

Z_2 and Z_3 sorry Z_1 and Z_3 all the way till covariance Z_1 and Z_n right similarly the off-diagonal values will be non 0 and will be driven by spatial dependence in data right?

And this spatial dependence is coming from you know the fact that this covariance will depend on this parameter theta through the parametric variogram model representation such that 2 gamma h is equal to 2 gamma h theta right and we know that theta is in a is itself in a parametric space capital theta.

Now, this 2 gamma h and 2 gamma h theta will provide me a representation of the covariance structure through the fact that 2 gamma h theta will be equal to 2 c 0 minus 2 c h, this is a theoretical relationship that we have studied earlier right?

So, that means, that the fact that I have a parametric variogram model is the fact that I can then formulate the covariance you know structure in the sigma theta matrix, and if I am able to then estimate theta, I can get to the variogram estimate as well because so, long as I have theta hat M L, I have my variogram 2 gamma hat which depends on the hat M L right?

So, what will really change in my analysis? What really will change is that I will now have the following log-likelihood function right because my variance-covariance matrix has changed from sigma squared i n to this you know sigma theta representation, it will provide me a different log-likelihood function. So, I have ln l which depends on mu and theta, theta can itself be a wet parameter right, sorry it can itself be a vector of multiple parameters and this is going to be minus n by 2 log of 2 pi minus 1 by 2 log of sigma theta.

So, the determinant of this matrix minus half Z_s minus mu transpose times sigma theta inverse, I am taking an inverse of this n by n matrix and I have Z_s minus mu. So, this is a little bit complicated well you know, I have a scalar entity on the left-hand side. So, on the left-hand side, I have a scalar entity the first entity is scalar the second entity is scalar I am working with a determinant value, but the last entity is more complicated it seems it is composed of vectors and matrices.

So, let us look at it more carefully. So, this Z, I know is a n by 1 vector mu itself is a n by 1 vector. So, this whole entity being transposed which is n by 1 minus n by 1 is going to be n by 1 transpose of n by 1 is going to be 1 by n sigma theta is n by n right its inverse of an n by n matrix is also n by n. So, I am going to have n by n in the middle part of this sandwich and the last part is going to be n by 1 minus n by 1 the whole thing will be n by 1.

Now, when we start to multiply these matrices we can look at the first two just to simplify our lives, the first two are 1 by n multiplied by n by n. So, the first thing to realize is they are conformable we can multiply them because the number of columns in Z_s minus mu is the same as the number of rows in sigma theta inverse right?

So, that is how we figure out whether or not the matrices are conformable and that is their product really defined. Here it is indeed defined and this product will turn out to be a 1 by n that is going to be a combination of this 1 here and this n here.

So, once I have this 1 by n matrix, I can go ahead and multiply it with this n by 1 matrix again these are conformable because the number of columns in the left-hand side matrix is the same as the number of rows in the right-hand side matrix, but the size of this matrix is going to be interestingly 1 by 1 which is nothing but a scalar. So, I really have a scalar entity on the right-hand side even though it might look a little bit more complicated than what we have seen till now right?

So, the fact is that I am trying to maximize my objective.

Again, the objective is to maximize because I am working in maximum likelihood. So, I am going to maximize log ln l mu theta let us just write it down quickly. So, that you know we have consistent notes for you ok. So, we have a scalar entity on the right-hand side, we want to maximize this log-likelihood by choosing mu and theta. So, our choice variables are mu and theta, and theta itself is composed of components like nugget and b C 0 and b for the linear model for linear variogram model right in case of the spherical model we had the nugget the sill and the range.

So, the number of choice variables will depend on which variogram model, do you want to work with as an analyst right whether you want to go with the linear or the spherical is your choice. So, I am going to say this is the analyst's choice, whichever way you go you can write down the first order conditions, solve them and you will have, I am going to say this exercise will yield theta hat M L right, that is in case of a linear variogram model we will have C 0 hat M L and we will have b hat M L right?

Once you have that you will also have a data-driven variogram 2 gamma hat h comma theta hat M L which is nothing but C 0 hat M L which is now a value data-driven understanding of spatial dependence which is exactly M L b hat M L and h. So, this is the data-driven representation of what is a nugget in my given sample and b hat is the rate of decline in spatial dependence as we move away from any given location by a lag vector h right?

So, we have a data-driven entity of b hat and C 0 hat M L. You should not feel terrified by these things one of the very reasons is that although these algorithms, when we work with them, they look mathematically intensive and you know sort of tedious most of these are going to be canned.

So, towards the end of this course, we will study you know we will do hands-on tutorials on r which will provide us you know the software syntax which will directly spit out the theta hat M L and you know the variogram estimator given the data set you have. But it is very important to know when you work on this software typically what happens is you run you write a code you run a syntax and a black box works you know which basically is maximizing all of these things and it is eventually giving the final answer, you know we may use, we will always use the final answer.

But we should know what is happening in that black box because when you work with real-world data sets when you work with complex problems when you do research you know you cannot really rely on what the software is giving you. If you do not understand what is happening behind the scenes in all likelihood you are going to get stuck at some point or your explanation of the physical world will come out to be nonsensical sometimes and so on and so forth right?

If you understand the machinery well great use the software right use the convenience of computation to do to actually calculate these things, but that does not discount the value of learning the process as it goes on you know mathematically ok. Now, there is a very important note that I want to give you before I end this maximum likelihood estimation is that M L estimates theta hat M L are biased right why?

They are biased because these do not account for degrees of freedom ok. These do not account for the degrees of freedom and what does it mean for us to say that, it does not account for degrees of freedom for which for this we will refer to the IID case where we had sigma hat squared M L equals summation i equals 1 to n Z_i minus Z bar the whole square divided by n.

Now, earlier in this course when I wrote you know sigma hat squared, I said it is rather n minus 1, and this n minus 1 account for a loss of degrees of one degree of freedom because in this definition I have already used Z bar; that means, although I am using Z_1 to Z_n by using also the knowledge of Z bar, I could have used any one of you know 1 less of this sequence of Z_1 to Z_n let us say I did not know Z_n.

I only knew Z_1 Z_2 all the way till Z_n minus 1 and I also know Z bar, I can back out Z_n; that means, in the definition of sigma hat squared M L, I am using the net variation of n minus 1 values given the knowledge of Z bar.

And hence in the denominator, I should have had an n minus 1 and not n.

But when you solve those first-order conditions in the IID case we can look at them now when we solve these you know 2 first-order conditions the sigma hat squared M L that comes out does not account for one the loss of a degree of freedom right? So, this loss of a degree of freedom, the loss remains unaccounted this is very important because it's one of its really fundamental knowledge that M L estimates are fundamentally biased for the variance-covariance estimators and the bias becomes larger and larger the smaller the data set that you are working with.

So, the next algorithm for achieving a fit of the variogram is called the least squares algorithm or the least squares estimator for a variogram model. Now as we move forward remember we started with experimental variograms right, what were experimental variograms well it's just 2 gamma h of j. So, the h j means h_1, h_2, h_3, h_4. So, different lag values in a given direction e right.

So, h_j is nothing, but the lag vectors in a given direction e right that is I have h_1, h_2, h_3 and keep going till h_n or let us say h let us not use n because n is the sample size let us say h_k ok. So, I have for the experimental variogram, I have manually gone in chosen k different lag values formed a vector of these lag values, and evaluated the variogram cloud for each of the hj lag values in a given direction let us say east-west direction and then you know did this for all different lag values.

So, k is the number of lags ok. Now the variogram model on the other hand will be given as 2 gamma h j e, but with a parameter vector theta. Now, theta we know is c 0 and b for a linear model right, and theta you know could be c 0 c s and a s for a spherical model right, and so on right? So, we have these different you know starting points. So, as an analyst, we should perhaps do more than one variogram model to see which one is providing a more sensible you know interpretation.

So, the least squares algorithm is as follows. We are trying to, first of all, we take a difference between 2 gamma h j which is the true value, which is the observed value of the data minus the mathematical model form which is either it could be linear or spherical or whatever I take this difference for every h.

Now, obviously, if I know the value of h the value 2 gamma h j theta will simply be a function of theta right? So, I can take this and take a square of it. So, it's a squared difference. So, the farther away you are from the given observed data I am going to penalize you by a square, this square difference is minimized because I want the least squares right?

So, I am squaring and then I want the least squares and that gives me a least squares algorithm for known h values the only variable in this entire formulation is theta. So, my choice variable when I am minimizing is going to be theta right? So, this exercise will give me, right? So, this will provide theta hat least-squares that are going to be let us say C o hat least squares and b hat least squares.

Now, remember C o hat least squares or b hat least squares are alternative estimates for the same model parameters nugget effect and the rate of decline in a spatial dependence through a different algorithm called the least squares algorithm. Earlier we evaluated C o hat or C 0 hat M L and b hat M L, we just have alternative estimates of the same here ok. Now I must specify that this is for a linear variogram, now the difference. So, the least squares algorithm has a few properties. So, we are just going to call them notes.

The first point is that the least squares algorithm provides a direct estimation of a variogram based on spatial dependence, right now this is different. So, I can say as opposed to the indirect M L estimation process procedure why was M L indirect? Because M L was first you know using the variance-covariance structure. So, it goes indirectly toward the variogram estimation, right?

So, indirect because it's through the variance-covariance matrix sigma theta right, but that does not make the maximum likelihood you know less legitimate in any way it is also going to be slightly biased.

But still, maximum likelihood is very tractable and it is you know also allows for very different you know it's much more general than least square something that you know you will realize as you work practically with data. The second point with least squares is that you know you can incorporate multiple directions can be incorporated by adding the respective squared differences.

So, if you add you know if you go back to this formulation you know, I mean you are saying that you just need to sort of add all the different squares right? So, here I have sort of done a short typo, I should add the least squares, right? So, I have squared the differences, but I should have added them from j equals 1 to k. So, if I add you know these more you know squared differences for let us say direction e tilde, I could simply just add them here and minimize the whole sum together for 2 different directions or 3 different directions and so on and so forth.

So, incorporating different directions is not so problematic or not so complicated with the direct estimation process of least squares as opposed to it what it would have been for the indirect you know maximum likelihood procedure ok.

The last topic that I want to touch upon in today's lecture is called the generalized least square squares algorithm. So, this sort of you know is sort of motivated by the fact that we have a sequence just like we have a sequence of data Z_1 to Z_n, and for this sequence of data we can write the mean and we can write the variance. Similar to that you know we have a sequence of these variogram values. So, we have this 2 gamma at h_1 2 gamma at h_2 and till 2 gamma till h you know some k value right?

So, we really have a vector of let us say k by 1 values of 2 gamma just like remember we had this vector for Z as well this was a n by 1 vector Z_1, Z_2, Z_n this. For this vector of data, we could write a mean value Z bar and we could also write a variance of Z bar right, we can simply calculate these things given the data set that we have. Similar to that you know there is going to be a variance of this 2 gamma value, I mean this 2 gamma is nothing but a vector.

So, I can just define it slightly differently here. So, I should be able to now figure out a variance-covariance structure of this value. Now, k by 1 if I have a vector which is k by 1 size the variance-covariance vector will be a k by k. So, I am really looking at the variance of 2 gamma h_1 variance of 2 gamma h_2. So, the variance of 2 gamma h_k, but then also covariance between the experimental plot at location 1 and location 2.

So, not only the data will exhibit spatial dependence, but in some way, even the variogram will exhibit that because you will have covariance with lag h_1 and covariance with lag h_2 and so on and so forth. So, if you have such a structure where there is a variance-covariance matrix such that you do not have you know the off-diagonal elements being 0 or all the diagonal elements being equal to each other then the least squares estimates are inconsistent and what you really need is a GLS algorithm right which looks like the following.

So, just like you know in the case of OLS we are going to minimize with choice variable theta, but now my formulation of what I am minimizing is going to be different. So, I am going to define this V as the variance-covariance matrix or I could say it's omega let us say ok. I have omega and omega inversed 2 gamma minus 2 gamma theta.

So, I have a slightly more complicated function to minimize what you will realize is if you look at the size of this you will see that it is also a scalar 1 by 1 let us do that right? 2 gamma by itself is a k by 1 2 gamma theta is going to be k by 1 the whole thing transpose is going to be 1 by k sigma inverse is k by k and this is just k by 1. The first two together are 1 by k and then multiplied by the last matrix is going to give me 1 by 1.

So, I am really minimizing a scalar value with respect to theta. I will write my first-order conditions, but this will yield theta hat GLS and not just LS, it's a generalized least squared estimate right, it is more robust, and it's a consistent estimator. So, and you know if you are interested in more information about this you can go back and read it from Cressy's book.

As far as getting to these values, are going to have all of this canned and software and we will do this in you know you can study this when you are looking at the software you know hands-on tutorial. But the point of the matter is that if you had a more complicated variance structure of the variogram the experimental variogram which is going to be likely right?

If the data themselves are spatially dependent on the second moment then the variogram is nothing but the representation of a second moment for intrinsically stationary data so; that means, that the second moment of this variogram representation is likely to be more complicated than what the least squares would you know assume that is the diagonal elements are the same and off-diagonal elements are 0 that is the least square assumption which may not be realistic.

What may be realistic is what we have seen on the slide in front of you, you must go over today's lecture one more time they are a little bit more involved and you must read Cressy’s book this material is in chapter 2 you must read it from the book as well.

So, thank you very much for your attention we will going forward, we will study you know we will look at variogram estimation in situations that are a little bit more interesting such that the spatial domain may be nonstationary or it might exhibit long-term trends which do not let these spatial domains to be stationary. They require a little bit more than what if we were to simply assume stationary domains and you know in the real world we may not always come across stationary domains.

So, we will go ahead and we will study that in the next lecture, and through that we will then introduce spatial prediction and study what is called spatial kriging ok.

So, thank you very much for your attention, and see you in the next lecture.