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

Lecture - 14B
Variogram Model Fitting

Welcome back, we are now going to study Variogram Model Fitting. So, the idea in variogram model fitting is to search for a valid variogram that is closest to the spatial dependence in a given sample data set.

So, what we have to start with is really a given sample data set. So, let us do that. So, we represent our sample data as a vector Z which is equal to Z at location S_1, Z at location S_2, and Z at location S_3, and keep going till location Z at location S_n right. Where location you know points S_1, S_2, till sn are my spatial entities at which I have sampled the data, everywhere else I have not sampled the data.

So, now you know so, this vector is all these values contained in this row vector transposed; that means, that Z by itself is a column vector. So, it is a vector of size n by 1 that is to say that Z really looks like an Excel you know a column in an Excel sheet where you have values spread out as you know along the cells in a column right? So, this is just to give you an idea of vector notation which is not really very difficult right?

So, next, you know let us consider the family the family of isotropic linear variograms right? So, we are working with isotropic linear variograms; that means, my lag vector h is diverse only to the extent of distance that it represents, in each direction the data that I am working with is going to exhibit the same spatial properties right. If the data were anisotropic then direction too would have become as important as distance right?

So, then if I were to do a North-South lag of 100 meters it will not give me the same spatial dependence structure as the East-West you know lag of 100 meters right? In our case when we study isotropic, direction really becomes you know a uniform representation of the process in all directions right? So, the direction is not really the entity of lag, it is really just a distance just to make my own life easier and also to restrict ourselves to the scope of this course right? We are not working with the isotropic data. So far this course is concerned.

So, we have seen this family previously it is called 2 gamma such that gamma h is equal to C 0 plus b h such that C 0 is greater than or equal to 0 and b is greater than equal to 0. With this family of linear variograms, our aim is to choose from the above family or set of variograms, the parametric model that is closest to the real-world situation or that is closest to the data that we have, that is we want the set p of 2 gamma such that 2 gamma h equals 2 gamma h and theta right?

And this theta lies in a definitive parametric space, it is just the space in which theta can move right? Theta cannot be arbitrarily anywhere it is supposed to be lying within a given set you know this capital theta ok.

So, for the linear variogram, this theta is really a vector by itself which contains C 0 and b, and ultimately the problem boils down to choosing parameters C 0 and b C 0 and b such that you know 2 gamma h theta fits right, fits well to the given sample data set ok. Now, you know we need to understand what does it mean? and what will it take to fit a good model to the given data set right?

So, for a fitting, the best model for fitting the best model on a variogram cloud, remember we always have this cloud from the experimental variogram right and we are trying to fit a model onto it, we will need an assumption on the distribution of spatial data right? So, as statisticians we view the world as a sequence of random processes, as spatial statisticians, we view the world as a sequence of random processes in space, right?

So, we need an assumption on the distribution right I mean. So, random processes will have specific distributions and we need to specify a distribution. So, what we need is for these values Z we need to be able to specify a CDF from which these values are being drawn, not only that the CDF representation F of Z will also provide us a property of how the values are connected in space right in a modeled sense.

Usually, F Z is assumed to be Gaussian, what does it mean when I say F is assumed to be Gaussian that is it is usually assumed to be normally distributed you know random variable or a random function, why? Well, because as we will see it, allows us to specify in nugget effect right?

We have seen that you know the nugget effect is usually considered white noise and that is why a Gaussian assumption on F of Z will allow us to specify a nugget effect, which is then very important to estimate an unbiased variogram, which then allows us to estimate an unbiased variogram ok. And the second thing that we will be needing to fit this best model is a “goodness of fit” criteria, I need a judgment device to figure out what is the best model or what is a better model relative to an alternative right? so, we need criteria.

In the case of regressions, I hope most of you have seen a regression model even if you have not, I am sure you must be aware of this statistic called R squared, R squared is a goodness of fit criteria right? There are multiple alternative goodness of fit criteria out there we will need such criteria to figure out what is the best model variogram for the real world ok.

So, we will now go over each algorithm that we have introduced. So, the first one was called the maximum likelihood estimation ok. Maximum likelihood estimation relies crucially on the Gaussian assumption on F of Z. Now, the process of maximum likelihood estimation aims to recover a parametric variogram model by exploiting the variance-covariance matrix of you know data in a spatial domain right?

So, let us write down the process of M.L.E, M.L.E which is an acronym for maximum likelihood estimation the process of M.L.E aims to exploit or use you know recover a variance-covariance matrix or structure of a spatial data set to provide an estimate of the variogram ok.

Now, to see this process let us begin with a case where the data are independent. So, with a no spatial dependence case right? So, we will start with the case which is simplistic and will have no spatial dependence. So, we will conduct MLE on it and then we will introduce spatial dependence to the structure. And then we will see how you know MLE will work to provide MLE an estimate of the parameter vector theta and hence a parameter model, variogram model you know 2 gamma h theta.

So, we are given a sequence of data, what is the sequence of data? It is Z_1, Z_2, Z_3, keep going all the way till Z_n ok. We define this as a vector Z ok again the data vector is a column vector. So, it is a n by 1 vector right? So, it is like you can imagine an Excel sheet and a column of cells where these data are recorded.

Now, remember we are given a sequence of data we have said that to do any analysis in maximum likelihood we need the Gaussian assumption right. So, it is the first crucial assumption that I must take that is we are going to say Z_i are iid that is independently and identically distributed as normal mu comma sigma squared. So, they are normally distributed with mean mu and variance sigma squared right.

Now, in this formulation I am going to do a little aside you know I could set if I set mu as let us say X_i beta where X_i is a vector of size 1 by k and beta is a vector of parameter k by 1 right? That is all I am writing here beta 1 x 1 i plus beta 2 x 2 i plus beta k x k_i then I am basically providing a setting that looks like a regression model right? So, then this setting would resemble the MLE or the maximum likelihood estimation of the following regression model which is Z_i equals beta 1 x 1 i beta 2 x 2 i beta 3 x 3 i plus u_i.

So, in this regression model, I have a systemic portion a systemic component, the systemic component is the one which can provide you a systematic you know how is the variation in Z_i explained by systematic factors x_1 till x_k. If Z_i were groundwater data x_1 might be rainfall, x_2 might be the amount of discharge, x_3 might be something like industrial demand, you know x_4 might be something like groundwater management policy which restricts the let us say number of electricity hours on a farm which draws groundwater and so on and so forth.

It could be the price of water, it could be alternative sources of water, it could be aquifer properties and so on and so forth. These are observed physical and social and economic factors around us that can explain how groundwater values would actually evolve over time or space right? These systemic components provide me a portion of the variation in Z_i that I as an analyst can explain by choosing $x_{i's}$ and then estimating betas in a linear coefficient form attached to each xi right? However, groundwater values are likely to be more complex in the way they evolve over space that is how the change by i.

The fact portion of the variation in $Z_{i's}$ over space that I cannot explain is captured in this residual or error component. This is the component of groundwater variation that I as an analyst cannot explain through this model right? The model is linear, it is restrictive, it could be highly non-linear right, one of the x_{is} x_{ks} could be squared or I could include them in square root. I could even multiply x_2 and x_3 and include that as a separate variable, I do not do any of that it is a simplistic model.

Again it is a model, it is an imperfect representation of the real world, but it provides me with a lot of information that is generalizable across space right. Because it is an imperfect representation there is a component, which remains you know as an error component and rather it is a random error component ok.

If I were to instead of mu if I were to say Z_i is iid N_{xi} beta I am basically talking about a regression model right? So, we will work with the sequence such that its mean is just mu, but you should be aware that we could just simply extend this formulation to a regression framework.

So, given coming back to our problem that Z_i is iid N mu comma sigma squared. The likelihood or probability likelihood of observing Z_i in the given sample is given as f which is the PDF of Z_i this is the continuous PDF representation and the parameters mu and sigma squared this f value is given as 1 over square root 2 pi sigma squared exponential minus half Z_i minus mu over sigma the whole squared.

The right-hand side mathematical formulation comes from the specification of a normal distribution, it represents a normal distribution, but given the value of Z_i once I know my Z_i value I can actually back out the probability or the propensity of observing this value in my sample what is the chance that I will observe let us say Z_i equals 5 given parameters mu and sigma squared right, this functional form will provide me that chance of observing each Z_i ok.

Now, after having learned the probability of observing a particular Z_i, I want to now figure out the likelihood or probability of observing the entire sample. So, if I have the probability of observing one Z_i one of the $Z_{i's}$ what is the probability of observing all the $Z_{i's}$ at once, the sequence Z_1, Z_2 through Z_n. Remember each Z_i that is Z_1, Z_2 is simply a random draw from this normal distribution.

So, the fact that I have a particular sequence of n values itself is an instance of chance right and I am trying to figure out what is the probability of observing the sample itself given that each Z_i is iid normal mu sigma squared. So, what I am after is f of Z_1, Z_2, Z_3 all the way till Z_n, I am looking for the probability of the sequence of data given mu and sigma squared right.

Remember I had said that these Z values put in a column vector can be defined as a vector Z of size N by 1. This Z is going to be distributed according to a multivariate normal of size mu and sigma squared, now because I have a Z vector of different $Z_{i's}$ each Z_i has an attached mean mu to it.

So, an N by 1 vector of $Z_{'s}$ will have a mean N by 1 attached to it, for variance a N by 1 vector will have a variance-covariance matrix of size n by n right. So, then sigma squared which is a scalar is not enough I need an identity matrix to be attached to it which is let us say sigma squared I n, I should be working with smaller you know n values. So, let me just make that correction in a minute n by 1 n by 1, and then n by n by multiplying sigma squared pi an identity matrix.

By doing that what I have really done is what I am saying is that the column vector Z_1, Z_2, Z_3 all the way to Z_n is distributed normally according to you know with a mean vector of mu mu mu of size n by 1 and a matrix n by n matrix having the diagonal values being sigma squared and off-diagonal values being 0 suggesting that the covariance between any of the $Z_{i's}$ is 0, which makes sense because you are working with no spatial dependence in data we are working with identically and independently observed data.

So, off-diagonal elements will be 0, when we introduce spatial dependence to the data we are exactly going to change the fact that the variance-covariance you know that is the off-diagonal elements of this covariance matrix is no longer going to be 0.

So, I have a n by 1 vector normally distributed iid with n by 1 mean vector and a n by n variance-covariance matrix, where the off-diagonal elements are 0 because I am working with the iid which is independently drawn distributions from the same you know which is the same distribution normal with the same mean and variance sigma squared across all $Z_{i's}$.

Now, when I write this when I get back to the problem that I am trying to solve is that I have to figure out what is the, I know f Z_i, I want f of Z vector which is all the values in my data sequence. This is going to be given as mu comma sigma squared I_n is equal to f of observing the first Z_1 times observing Z_1 conditional on sorry observing Z_2 condition on the fact that I have observed Z_1 times, I have the probability of observing Z_3 conditional on the fact that I have observed Z_2 and Z_1 keep going all the way till observing Z_n condition on the fact that I have observed Z_n minus 1, Z_n minus 2, all the way till Z_1 ok.

Now, the fact that we have an iid assumption, so, given the iid assumption what happens is that the conditional density is equal to the marginal density. That is the probability of observing Z_2 does not depend on what Z_1 was or the probability of observing Z_n does not depend on what happened earlier in when we draw you know Z_1, Z_2 all the way till Z_n minus 1. It is independent of what is happening before or after it that is not going to be the case if we had spatial dependence in data.

But with the iid assumption we can write f of Z mu comma sigma squared In equals f of Z_1 times f of Z_2 times f of Z_3 times f of Z_4 times keep going till f of Z_n. That is we have a multiplicative representation i that goes from 1 to n f of Z_i which is to say we have a multiplication of i that goes from 1 to n1 over square root 2 pi sigma squared exponential minus half Z_i minus mu over sigma the whole squared.

And this f of Z_i, I am sourcing directly from what I have you know presented earlier. If I were to solve this, I am simply going to get, I am going to multiply each component n times. So, I am going to get 1 over 2 pi sigma square to the power n over 2 times exponential. So, you have exponentials e to the power you know this stuff is multiplied for each Z_i. So, I am going to write this is going to be minus half summation, I go 1 to n Z_i minus mu by sigma the whole squared.

Let us take it to the next page. So, we have the likelihood of observing our sample given Z_i are iid normal 0 sigma squared being given as of f of Z vector which is a n by 1 vector given mu and sigma squared equal to 1 over 2 pi sigma square to the power n by 2 exponential summations minus, half i equals 1 to n Z_i minus mu sigma the whole squared alright.

Now, with that you know we take a log of this and that is called the log-likelihood. So, the log-likelihood of observing the given sample so, I can call this f, I can define this as l of you know mu and sigma squared the 2 parameters that I am after log-likelihood of this is going to be l n l mu comma sigma squared equals minus n by 2 log of 2 pi minus n by 2 log of sigma squared minus 1 over 2 sigma squared summation i equals 1 to n Z_i minus mu the whole squared.

Now, my objective is to maximize. So, my algorithm is called maximum likelihood estimation I have the likelihood of observing the sample. The next step is to maximize the likelihood of observing this sample. So, I am just going by the nomenclature I am going to maximize this, and when I maximize this my choice variables are mu and sigma square. So, these are the variables that I can choose, I have a degree of freedom about what mu and sigma squared values could be right?

So, I could choose mu and sigma squared such that the likelihood of observing this given sample the sequence Z_1 till Z_n is maximized given that $Z_{i's}$ are iid normally distributed with parameters mu and sigma squared. So, this is an optimization problem and when we face such an optimization problem we write our first-order conditions. So, our first order conditions will be del l n l by del mu which is the first partial differential by our first choice variable this will be set equal to 0, then del l n l by del sigma squared which is our second choice variable this will be set equal to 0.

Overall I have 2 equations and 2 unknowns ok. So, I can solve this. When I actually solve this we will find that mu hat ML is the maximum likelihood estimator for mu will come out to be Z bar which is nothing but the sample mean of all the $Z_{i's}$ whereas, sigma hat squared ML which is where the hat means it is a data-driven estimate and ML means it is a maximum likelihood data-driven estimate, that is going to come out to be summation i equals 1 to n Z_i minus mu hat ML which is nothing but the Z bar value divided by n ok.

Now, these are my maximum likelihood estimator's maximum likelihood estimators without spatial dependence ok. So, as a next natural step, what we are going to do is that we are going to add spatial dependence to these data and see how what happens and how these parameter estimates change what changes when we add spatial dependence to the data. The first thing that is going to happen is that I am not going to have sigma squared I_n as my variance-covariance matrix.

My variance-covariance matrix will be more complicated because my off-diagonal elements are going to be non-zero. So, the first change that you are going to observe is at the off-diagonal elements of the variance-covariance matrix of this vector. So, let us move forward and do that.