**Lecture 11 Spatial Correlation, Variogram and Models**

In spatial data, observations close together in space will probably look more similar to one another than observations collected farther away from one another, so when we fit any kind of trend surface to these data, the errors from the model may be correlated. In this talk we discuss tools for modelling spatial correlation.

In geosta tistics the spatial correlation is modelled by the ***variogram***. Here, the word variogram will be used synonymously with ***semivariogram*** (there are some different definitions in some texts). The variogram plots semivariance as a function of distance.

To estimate the spatial correlation from observational data, we need to make ***stationarity*** assumptions. One commonly used form of stationarity is *intrinsic* (***weak***) stationarity.

**11.1 Spatial correlation**

In Lecture 6 we defined the lag-*k* covariance and correlation. These measure the covariance and correlation between two observations that are *k* units apart in time for a stationary time series, i.e., the time series has a constant mean and the covariance depends only on the lag *k*.

Extending this idea to spatial data, we might assume that the observations represent a snapshot from a random process over space. Weak stationarity implies that the surface has a constant mean and that the covariance between two observations depends only on the distance (and perhaps direction) between the locations of these observations. The covariance is ***isotropic*** if it depends only on the distance between the locations and not on the direction:

Cov[Z(), Z()] = *C*(h)

where *h* is the distance between location  and location . The function *C* is called the ***covariogram*** and is analogous to the auto covariance function for time series data. The function

ρ(h) = C(h)/C(0)

is called the ***correlogram*** and is analogous to the autocorrelation function for time series data. Rather than looking at covariograms and correlograms, people who deal with spatial data often use the variogram, which is defined as:

 = Var[Z() - Z()]

= Var[Z()] - Cov[Z(), Z()]

= C(0) – C(h)

The estimator of  is given by

 = 

where *N*(*h*) denotes the set of all pairs of locations that are *h* units apart, |*N*(*h*)| denotes the number of pairs in this set. There are three features in variograms:

* Sill – The sill is equal to the variance of the process, i.e., the covariance *C*(h) at distance *h* = 0;
* Range – The range is the distance at which observations are no longer correlated. The range may be finite or infinite;
* Nugget effect – If ≠ 0. The nugget effect represents micro-scale variation and/or measurement error.

**Example 11.1** Empirical variogram for the benthic data:

|  |
| --- |
| library(sp)  library(EnvStats)  library(gstat)  data(Benthic.df)  coordinates(Benthic.df)=~Longitude+Latitude  vg.benthic <- variogram(Index ~ 1, data=Benthic.df)  plot(vg.benthic, main="Figure 11.1 Empirical variogram for Benthic Index") |
|  |

There appearssome nugget effect.  **█**

**11.2 Variogram modelling**

Several models have been postulated for isotropic spatial correlation. Three common variogram models are the exponential, Gaussian, and spherical:

Exponential: C(h) = 

Gaussian: C(h) = 

Spherical: C(h) = , *h* < r

where  is the variance of the process and *r* is a constant. Note that the range for the exponential and Gaussian models is infinite, while for the spherical model it is equal to the constant *r*.

**Example 11.2** Fitting variogram models to the benthic data:

|  |
| --- |
| library(sp)  library(EnvStats)  library(gstat)  data(Benthic.df)  coordinates(Benthic.df)=~Longitude+Latitude  vg.benthic <- variogram(Index ~ 1, data=Benthic.df)  vg.fit.benthic<-fit.variogram(vg.benthic, model=vgm(1,"Exp", 0.5,1))  plot(vg.benthic, vg.fit.benthic,main="Figure 11.2 Exponential variogram for Benthic Index")  vg.fit.benthic<-fit.variogram(vg.benthic, model=vgm(1,"Gau", 0.5,1))  plot(vg.benthic, vg.fit.benthic,main="Figure 11.3 Gaussian variogram for Benthic Index")  vg.fit.benthic<-fit.variogram(vg.benthic, model=vgm(1,"Sph", 0.5,1))  plot(vg.benthic, vg.fit.benthic,main="Figure 11.4 Spherical variogram for Benthic Index") |
|  |

█

**11.3 Directional variograms**

We have assumed that the spatial process is stationary and that the spatial correlation is isotropic. We can use directional variograms explore whether these assumptions appear to be valid for the data. Example 11.3 displays the empirical semivariogram for the benthic index for four difference directions.

**Example 11.3** Directional variograms for Benthic Index:

|  |
| --- |
| library(sp)  library(EnvStats)  library(gstat)  data(Benthic.df)  coordinates(Benthic.df)=~Longitude+Latitude  dvg.benthic <- variogram(Index ~ 1, data=Benthic.df, alpha=c(0, 45,90,135))  dvg.fit.benthic<- vgm(1,"Sph", 0.5,1,anis=c(30,0.4))  plot(dvg.benthic,dvg.fit.benthic, main="Figure 11.5 Directional variograms for Benthic Index") |
|  |

Note: zero direction is North; 90 degrees is East; and so on.

The form of the variogram appears to be different for different directions based on the raw benthic index data. Differences in the variogram in different directions may be caused by the presence of trend and/or anisotropy (different forms of spatial correlation in different directions). █

**Exercises**

11.1 Repeat the examples in this talk.

11.2 Re Example 11.3, obtain variograms with four directions: 22.5, 67.5, 112.5 and 157.5.

**References**

* Bivand, R. S., Pebesma, E. and Gómez-Rubio, V., (2013), *Applied Spatial Data Analysis with R*, SPRINGER
* Millard, S.P. and Neerchal, N. K. (2000), *Environmental Statistics with S-PLUS*, Chapman & Hall.