Simple Overview of Statistical Concepts of week 6

Ordinary Least Squares Fit. Ordinary least squares (OLS) is a statistical method used to find the curve of best fit that explains the relationship between two or more variables. It is often used in regression analysis to fit a linear or non-linear curve to a set of points. The method works by minimizing the sum of the squared errors between the observed values and the predicted values obtained from the theoretical model. In spatial data analysis it is often used to fit a theoretical variogram to its empirical estimate. Usually, a suitable theoretical variogram can be selected by visual inspection of a sample variogram obtained from spatial data. It is often assumed that parameters of such selected theoretical variogram are unknown. OLS estimate these parameters to guarantee the closes fit of theoretical variogam to the sample one.

Maximum Likelihood Methods Maximum likelihood methods are used to estimate the parameters of a statistical model by maximizing the likelihood function. The likelihood function is a probability distribution function that measures the likelihood of observing the data given a set of parameters. Maximum likelihood estimation (MLE) is used to find the values of the parameters that maximize the likelihood of the data, assuming a specific probability distribution. Similar to OLS, the parameters estimated by MLE guarantee the closes fit of theoretical variogam to the sample data, but in another sense than OLS. Instead of distances between points, MLE approaches are concentrated on the likelihood of the sampled observations.

Initial values for variogram fit. To estimate a variogram, an initial guess of its parameters must be provided, which is known as selecting initial values for the variogram fit. These initial values are subsequently improved by numeric methods and should converge to the true values of parameters of the variogram. This step is important because a poor initial guess can result in a poorly fitting variogram or errors. It is essential to choose appropriate initial values for the parameters in order to ensure convergence. The initial values have a significant impact on the performance of the algorithm, and choosing suitable values can help to speed up the convergence process and increase the accuracy of the solution.

There are several approaches to selecting initial values for the variogram fit. One is to use prior knowledge of the study area to inform the initial guess. Another approach is to visually inspect the data and roughly estimate the range, nugget and other parameters of the theoretical variogram from the sample variogram. Often, to select the initial values one uses a grid search. Grid search involves selecting a set of possible initial values for the parameters that form a grid over a parameter space and then systematically testing each combination of values to determine which yields the best result. This can be a time-consuming process.

Residual variogram. A residual variogram is used to analyse the spatial dependence between measured values of a variable at different locations. It is computed by subtracting the estimated spatial trend from the observed data and computing sample variogram for the obtained residuals. The residual variogram can help to estimate patterns in the spatial dependence that are not captured by the chosen spatial trend model. It also can help to identify areas where the model may need to be improved.

Anisotropy. In some cases, the relationships between spatial locations may depend on the direction or orientation of the data. In such cases, the sample covariance may be anisotropic, meaning that it varies depending on the direction or orientation of the data. When the data are isotropic, for a selected location the dependencies are same with all locations on a circle around it. However, when the covariance is anisotropic, these circles are stretched and transformed to ellipses. The amount of stretching in each direction depends on the strength of dependencies along that direction.

Properties of covariance functions. Not all functions are covariance functions. If a function fails to satisfy the known properties of covariance functions, it cannot be considered a valid covariance function. On the other hand, if a function satisfies these properties, it does not necessarily guarantee that it is a valid covariance function. There may be other properties that it fails to satisfy, which can impact its validity as a covariance model. Therefore, there are several methods that are useful to obtain basic covariance functions, which can be used to build more complex covariance models.