Ordinary Kriging in Dialectology

Jack Grieve
Aston University
j.grieve1@aston.ac.uk

There are many situations that arise in dialectology where it is useful to be able to estimate the values of a linguistic variable at one or more unobserved locations. The standard approach to spatial interpolation in geography and other fields is known as ordinary kriging (Isaaks and Srivastava, 1989), although ordinary kriging in particular and interpolation in general has rarely been applied in dialectology (although see Grieve, 2013). This presentation will therefore introduce ordinary kriging and demonstrate how this geostatistical technique can be applied in dialectology by presenting a series of studies that focus on regional variation in American English. In addition, the presentation will introduce the implementation of these techniques in R (see Bivand et al, 2008).

Ordinary kriging is a method for interpolating the value of a variable at an unobserved location based on the values of that variable at observed locations. Specifically, ordinary kriging estimates the value of a variable at an unobserved location by taking a weighted average of the values of the variable at observed locations, where these weights are based on both the distance separating the locations and the variogram for that variable, which is a function that describes the amount of spatial variability in the values of a variable measured over a series of locations (Isaaks and Srivastava, 1989). A variogram plots the variance between locations against the distance between locations, in essence providing a model of how the values of a variable change across space. Once a variogram has been estimated for a variable based on the values of that variable at observed locations, ordinary kriging can then be used to estimate the values of that variable at unobserved locations.

After introducing variogram analysis and ordinary kriging, a variety of different applications of ordinary kriging in dialectology will be demonstrated through analyses of regional linguistic variation in American English. In particular, the use of ordinary kriging for facilitating prediction, visualization, comparison, and aggregation of regional linguistic data will be discussed, based on a variety of American English datasets, including phonetic data from the Atlas of North American English (Labov et al., 2006), grammatical data from a corpus of written American English (Grieve et al., 2011), and lexical and phonological data from the Harvard Dialect Survey (Vaux, 2003).

First, the basic use of ordinary kriging to predict the values of linguistic variables at unobserved locations will be demonstrated through an analysis of a variety of individual linguistic variables.

Second, the use of ordinary kriging to estimate the values of a linguistic variable across an entire region at a very high level of resolution will be discussed. This is an especially powerful method for visualizing dialect data, as it allows for general patterns of regional variation to be mapped based on only the values of a relatively small number of known locations. For example, Figure 1 plots the values of an aggregated phonetic variable that was originally measured across 236 observation points after being interpolated across approximately 20,000 regularly spaced locations using ordinary kriging. In this way ordinary kriging can be used to plot isoglosses, dividing a region into sub-regions where the different values of the variable predominate.
Third, the use of ordinary kriging to facilitate the comparison of dialect maps that are based on different sets of locations will be demonstrated by interpolating maps from different dialect surveys over a regular grid of reference locations. These interpolated maps will then be correlated with each other in order to measure the similarity between the regional patterns identified in these various dialect surveys.

Finally, the use of ordinary kriging to facilitate the aggregation of dialect maps that are based on different sets of locations will be demonstrated by aggregating linguistic variables from different American dialect studies after each of the individual linguistic variables have first been interpolated over a regular grid of reference locations.

References


