

Shahin Tavakoli
Department of Statistics, University of Warwick, UK

Abstract

Title: Spatial modeling approach to dialect sound variation

A Spatial Modeling Approach for Linguistic Object Data: Analysing dialect sound variations across Great Britain Dialect variation is of considerable interest in linguistics and other social sciences. However, traditionally it has been studied using proxies (transcriptions) rather than acoustic recordings directly. We introduce novel statistical techniques to analyse geolocalised speech recordings and to explore the spatial variation of pronunciations continuously over the region of interest, as opposed to traditional isoglosses, which provide a discrete partition of the region. Data of this type require an explicit modeling of the variation in the mean and the covariance. Usual Euclidean metrics are not appropriate, and we therefore introduce the concept of d -covariance, which allows consistent estimation both in space and at individual locations. We then propose spatial smoothing for these objects which accounts for the possibly non convex geometry of the domain of interest. We apply the proposed method to data from the spoken part of the British National Corpus, deposited at the British Library, London, and we produce maps of the dialect variation over Great Britain. In addition, the methods allow for acoustic reconstruction across the domain of interest, allowing researchers to listen to the statistical analysis.

This is joint work with Davide Pigoli and John Aston (Cambridge), and John Coleman (Oxford).