Statistical Sciences
UNIVERSITY OF TORONTO

SEMINAR

February 13, 2017 at 3:30 pm

Refreshments will be provided at 3:15pm

Innis College, Room 204

Speaker: Shahin Tavakoli, University of Cambridge

Host: Fang Yao

Functional Data Analysis driven by modern applications

Functional Data Analysis is an instantiation of a modern shift in modeling paradigm: it acknowledges the fact that the data discretization is an artifact of the measurement process, a simplification of the reality of the generative process giving rise to the data. As such, it is an invitation to revisit and expand existing (multivariate) methodologies.

I will present an overview of my previous work, starting by a framework for frequency domain analysis of functional time series, with some basic theory, and an application to molecular dynamics data. I will then follow with some theoretical work on the covariance separability assumption for multidimensional functional data, which is often implicitly used in many neuroimaging applications.

The bulk of my talk will be on a recent project on spatial modeling of dialect variations, using acoustic speech recording.

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 geo-localised 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. Although the methodology is applied to sounds data, some of it (particularly the d-covariance) could be applied to neuroimaging data, such as diffusion tensor imaging.