Functional Clustering Methods and Marital Fertility Modelling

Per Arnqvist

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för avläggande av filosofie doktorsexamen framläggs till offentligt försvar i Hörsal 1, Lindellhallen,
Fredagen den 24 februari, kl. 10:15.
Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent: Professor Helle Sørensen,
Department of Mathematical Sciences, University of Copenhagen, Denmark.

Author

Per Arnqvist

Title

Functional Clustering Methods and Marital Fertility Modelling

Abstract

This thesis consists of two parts.
The first part considers further development of a model used for marital fertility, the Coale-Trussell's fertility model, which is based on age-specific fertility rates. A new model is suggested using individual fertility data and a waiting time after pregnancies. The model is named the waiting model and can be understood as an alternating renewal process with age-specific intensities. Due to the complicated form of the waiting model and the way data is presented, as given in the United Nation Demographic Year Book 1965, a normal approximation is suggested together with a normal approximation of the mean and variance of the number of births per summarized interval. A further refinement of the model was then introduced to allow for left truncated and censored individual data, summarized as table data. The waiting model suggested gives better understanding of marital fertility and by a simulation study it is shown that the waiting model outperforms the Coale-Trussell model when it comes to estimating the fertility intensity and to predict the mean and variance of the number of births for a population.

The second part of the thesis focus on developing functional clustering methods.
The methods are motivated by and applied to varved (annually laminated) sediment data from lake Kassj\"on in northern Sweden. The rich but complex information (with respect to climate) in the varves, including the shapes of the seasonal patterns, the varying varve thickness, and the non-linear sediment accumulation rates makes it non-trivial to cluster the varves. Functional representations, smoothing and alignment are functional data tools used to make the seasonal patterns comparable.
Functional clustering is used to group the seasonal patterns into different types, which can be associated with different weather conditions.

A new non-parametric functional clustering method is suggested, the Bagging Voronoi K-mediod Alignment algorithm, (BVKMA), which simultaneously clusters and aligns spatially dependent curves. BVKMA is used on the varved lake sediment, to infer on climate, defined as frequencies of different weather types, over longer time periods.

Furthermore, a functional model-based clustering method is proposed that clusters subjects for which both functional data and covariates are observed, allowing different covariance structures in the different clusters. The model extends a model-based functional clustering method proposed by James & Sugar (2003). An EM algorithm is derived to estimate the parameters of the model.

The resulting clusters from the different functional clustering methods and their time dynamics show great potential for seasonal climate interpretation, in particular for winter climate changes.