Model-based Clustering of Multivariate Longitudinal Data of a Mixed Type

Jan Vávra

October 3, 2022

Abstract

In many nowadays studies, the data are collected repeatedly on the same units over a certain period of time. Moreover, such longitudinal data are composed of numeric values, count variables, binary indicators, ordered or nominal categories. A few variants of statistical model capable of modelling such often highly correlated data jointly are introduced. On top of that, a methodology of model-based clustering is adapted to such models to discover hidden heterogeneity within the data by dividing units into clusters of specific characteristics. Bayesian approach is taken, generative model is proposed and MCMC methodology is developed for estimation. A simulation study verifying the estimation properties is conducted. The methodology is applied to real datasets such as medical data on patients suffering from primary biliary cholangitis (PBC) or economical dataset consisting of thousands of Czech households followed since 2005 (EU-SILC database).