Clustering of longitudinal data based on a random change-point model using Dirichlet processes

Elaheh Shams[1*] Reyhaneh Rikhtehgaran[1]
[1]Department of Mathematical Sciences, Isfahan University of Technology

Abstract

A dynamic regression model which switches between two models with different slopes at random time points is considered. Model-based clustering of longitudinal data is conducted by assuming the Dirichlet process (DP) as a prior for the distribution of random change points. The discreteness nature of the DP is utilized to cluster subjects according to the time of changing slopes. Markov chain Monte Carlo simulation methods are adopted to achieve the parameter estimates. Performance of the proposed model is illustrated by conducting a simulation study.

Keywords: Bayesian approach, Change-point model, Dirichlet processes, Gibbs sampler.
Mathematics Subject Classification [2010]: 62M99, 62J05, 62H30

1 Introduction

In recent years increasing interest has been shown in the problem of clustering of longitudinal data. These data sets are available by repeatedly measuring subjects through time. Usually in Econometrics and Biostatistics, it frequently happens that effect of occurring an event on changing slope of response variable against time will appear in different times lag for different subjects. Thus, it is usually important that different subjects can be categorized based on these change-point times. For example, when an economical event leads to profitability declines for companies in the stock market, different firms show different tolerances such that face values of some companies start decreasing sooner than others. Thus, clustering of firms based on the time of standing against the decline is useful. Addressing this issue, we propose the use of the Dirichlet process (DP) in the structure of a dynamic change-point model.

The rest of this paper is organized as follows. In Sections 2, we briefly introduce the DP prior. Section 3 specifies the proposed dynamic change-point model. In Section 4, we implement the Gibbs sampling scheme. The last section presents a simulation study.

*Speaker