Change point detection problem
in time series data

Daniel Kucharczyk

Abstract

Real world signals are modelled using time series methods, in order to extract meaningful information about given phenomena. Regularly, these observations reveal properties associated with non-stationary processes. However, non-stationary processes are difficult to analyse using classical statistical tools, because the majority of the tools have been designed to work only with stationary processes (processes having homogeneous properties).

Consequently, in order to gain insight about a phenomenon that is in doubt, there is a need to decompose the associated non-stationary time series data into sets of homogeneous segments of data. This type of decomposition may be achieved using a change point analysis. In this type of analysis, the way in which a given sequence changes, at defined points in time, is estimated and properly assessed. The obtained estimates of positions (points in time) are known as change points.

In this thesis, several novel change point detection techniques used for both simulated and real-time series data have been indicated. Moreover, one focus of this work is how to deal with the problem of detecting change points in the scale (or variance) of a time series process. A review of the literature reveals that a wide spectrum of mathematical techniques (case-by-case basis) is required to efficiently tackle the initial problem. Herein, special attention has been paid to the case of time series data driven by processes with the property of long memory. This is the phenomenon whereby events which are arbitrarily distant still influence each other.

In the fist part of the thesis, a general introduction to the theory of change point analysis is presented and it is accompanied by an extensive literature review.

In the next step we consider problems associated with detecting change points for time series data with independent observations. We analyse the case of Gaussian processes. This focuses on the case when a non-stationary time series is driven by a mixture of processes, with at least one being a Gaussian process. The discussion about Gaussian processes begins with the presentation of a simple but very powerful technique based on the idea of applying

Gaussian filters to initial data. The application of a Markov regime-switching model, as applied to the change point detection problem, is also presented as an alternative statistical tool.

In addition we consider the problem of variance change point detection for independent observations without making any particular assumption about data distribution. Two different methods are presented. Specifically, a new change point detection method based on the absolute deviation about the median in conjunction with Adaptive Regression Splines technique is presented. Furthermore, an alternative approach for detecting multiple variance change points by taking into consideration locally defined test statistics is presented.

Moreover, we present the problem of detecting change points in time series data where it can be assumed that the observations are no longer independent; in particular, change point problems related to the shifts of variance for dependent observations as a natural extension of a classical Gaussian model are discussed. In this model the observations do not need to be independent. To be more precise, the problem of variance change point detection for fractional Gaussian noise based on the likelihood ratio procedure is presented. Also, an alternative approach based on the cumulative sums of squares test is developed and applied to fractional Gaussian noise.

Along with the theoretical results, the extensive simulation studies is performed. A comparison of the application of selected methods to real problems and results obtained using competitive models is also presented.

In the last part of the thesis, brief conclusions with the summary of main results are presented.