Statistica Sinica **37** (2027), 1-21
doi:https://doi.org/10.5705/ss.202024.0182

SEQUENTIAL CHANGE POINT DETECTION IN HIGH-DIMENSIONAL VECTOR AUTO-REGRESSIVE MODELS

Yuhan Tian and Abolfazl Safikhani*

University of Florida and George Mason University

Abstract: Sequential/Online change point detection involves continuously monitoring time series data and triggering an alarm when shifts in the data distribution are detected. We propose an algorithm for real-time identification of alterations in the transition matrices of high-dimensional vector auto-regressive models. This algorithm initially estimates transition matrices and error term variances using regularization techniques applied to training data, then computes a specific test statistic to detect changes in transition matrices as new data batches arrive. We establish the asymptotic normality of the test statistic under the scenario of no change points, subject to mild conditions. An alarm is raised when the calculated test statistic exceeds a predefined quantile of the standard normal distribution. We demonstrate that as the size of the change (jump size) increases, the test's power approaches one. Empirical validation of the algorithm's effectiveness is conducted across various simulation scenarios. Finally, we discuss two applications of the proposed methodology: analyzing shocks within S&P 500 data and detecting the timing of seizures in EEG data.

Key words and phrases: Auto correlation, break point, sequential data, structural break, temporal dependence.

1. Introduction

Abrupt changes in daily life are often perceived as anomalies, typically requiring careful study to prevent future repercussions. In a data set, such abrupt changes are usually triggered by shifts in the data-generating process. Detecting these changes precisely and quickly is essential for understanding their origins and mitigating potential harm. Consequently, change point detection (CPD) has become a critical research area in both data science and statistics, with real-world applications spanning power systems, quality control, and advertising (Routtenberg and Xie, 2017; Page, 1954; Zhang et al., 2017). Most CPD methods, classified here as offline CPD, require access to the full dataset and aim to pinpoint change point locations accurately. However, with advancements in cloud storage and computing, streaming data has become ubiquitous, necessitating a different CPD approach for monitoring incomplete and dynamic data in real

*Corresponding author. E-mail: asafikha@gmu.edu