Statistica Sinica **35** (2025), 1-25
doi:https://doi.org/10.5705/ss.202022.0323

INFERENCE FOR CHANGE POINTS IN HIGH DIMENSIONAL MEAN SHIFT MODELS

Abhishek Kaul* and George Michailidis

Washington State University and University of California, Los Angeles

Abstract: We consider the problem of constructing confidence intervals for the locations of change points in a high-dimensional mean shift model. We develop a locally refitted least squares estimator and obtain component-wise and simultaneous rates of estimation of change points. The simultaneous rate is the sharpest available by at least a factor of $\log p$, while the component-wise one is optimal. These results enable existence of limiting distributions for the locations of the change points. Subsequently, component-wise distributions are characterized under both vanishing and non-vanishing jump size regimes, while joint distributions of change point estimates are characterized under the latter regime, which also yields asymptotic independence of these estimates. We provide the relationship between these distributions, which allows construction of regime adaptive confidence intervals. All results are established under a high dimensional scaling, in the presence of diverging number of change points. They are illustrated on synthetic data and on sensor measurements from smartphones for activity recognition.

Key words and phrases: High dimensional, inference, limiting distributions, multiple change points, optimal estimation, regime adaptation.

1. Introduction

Statistical models with multiple change points are of significant interest due to their numerous applications in diverse areas, including economics and finance (Frisén, 2008), functional genomics and neuroscience (Koepcke, Ashida and Kretzberg, 2016) amongst others. These have been studied extensively for a variety of statistical models, including mean shifts, regression, graphical models, factor and specific time series models and various algorithms have been developed to accomplish this task -dynamic programming, regularized cost functions, multiscale methods, etc. See, e.g., the review by Niu, Hao and Zhang (2016).

The main statistical tasks in change point analysis aim to address the following questions: (i) *whether change point(s) exist* in the data, (ii) assuming their existence, *estimation* of their location and (iii) *post-estimation inference*. The literature on the first two aspects is extensive and a multitude of methods for a variety of models is available, including under a fixed dimension p setting

*Corresponding author. E-mail: abhishek.kaul@wsu.edu