High-dimensional Change-point Detection Using Generalized Homogeneity Metrics

Shubhadeep Chakraborty[1] and Xianyang Zhang[2]

[1]University of Washington
[2]Texas A&M University

Abstract

Change-point detection has been a classical problem in statistics and econometrics. This work focuses on the problem of detecting abrupt distributional changes in the data-generating distribution of a sequence of high-dimensional observations, beyond the first two moments. This has remained a substantially less explored problem in the existing literature, especially in the high-dimensional context, compared to detecting changes in the mean or the covariance structure. We develop a nonparametric methodology to (i) detect an unknown number of change-points in an independent sequence of high-dimensional observations and (ii) test for the significance of the estimated change-point locations. Our approach essentially rests upon nonparametric tests for the homogeneity of two high-dimensional distributions. We construct a single change-point location estimator via defining a cumulative sum process in an embedded Hilbert space. As the key theoretical innovation, we rigorously derive its limiting distribution under the high dimension medium sample size (HDMSS) framework. Subsequently we combine our statistic with the idea of wild binary segmentation to recursively estimate and test for multiple change-point locations. The superior performance of our methodology compared to other existing procedures is illustrated via extensive simulation studies as well as over stock prices data observed during the period of the Great Recession in the United States.

Keywords: High Dimensionality, Multiple Change-Point Detection, Two Sample Test, Wild Binary Segmentation.