High-dimensional changepoint estimation with heterogeneous missingness

Tengyao Wang

The London School of Economics and Political Science

Abstract:

We propose a new method for changepoint estimation in partially-observed, high-dimensional time series that undergo a simultaneous change in mean in a sparse subset of coordinates. Our first methodological contribution is to introduce a 'MissCUSUM' transformation, that captures the interaction between the signal strength and the level of missingness in each coordinate. In order to borrow strength across the coordinates, we project these MissCUSUM statistics along a direction found as the solution to an optimisation problem. The changepoint can then be estimated as the location of the peak of the projected series. In a model that allows different missingness probabilities in different component series, we identify that the key interaction between the missingness and the signal is an observation-probability-weighted sum of squares of the signal change in each coordinate. More specifically, we prove that the angle between the estimated and oracle projection directions, as well as the changepoint location error, are controlled with high probability by the sum of two terms, both involving this weighted sum of squares, and representing the error incurred due to noise and due to missingness respectively. The striking effectiveness of our methodology is further demonstrated both on simulated data, and on an oceanographic data set.