OPTIMAL CHANGE-POINT DETECTION AND LOCALIZATION

NICOLAS VERZELEN

Abstract: Given a times series Y in $?^n$, with a piecewise constant mean and independent components, the twin problems of change-point detection and change-point localization respectively amount to detecting the existence of times where the mean varies and estimating the positions of those change-points. In this work, we tightly characterize optimal rates for both problems and uncover the phase transition phenomenon from a global testing problem to a local estimation problem. Introducing a suitable definition of the energy of a change-point, we first establish in the single change-point setting that the optimal detection threshold is $\sqrt{2\log\log(n)}$. When the energy is just above the detection threshold, then the problem of localizing the change-point becomes purely parametric: it only depends on the difference in means and not on the position of the change-point anymore. Interestingly, for most change-point positions, it is possible to detect and localize them at a much smaller energy level. In the multiple change-point setting, we establish the energy detection threshold and show similarly that the optimal localization error of a specific change-point becomes purely parametric. Along the way, tight optimal rates for Hausdorff and l^1 estimation losses of the vector of all change-points positions are also established. Two procedures achieving these optimal rates are introduced. The first one is a least-squares estimator with a new multiscale penalty that favours well spread change-points. The second one is a two-step multiscale post-processing procedure whose computational complexity can be as low as O(nlog(n)). Notably, these two procedures accommodate with the presence of possibly many low-energy and therefore undetectable change-points and are still able to detect and localize high-energy change-points even with the presence of those nuisance parameters.