Perturbation-based confidence regions and their application to stochastic bandits

Master's thesis - Abstract

Author: Botond Kiss

Supervisors: Dr. Balázs Csanád Csáji, Dr. Gábor Pete

Constructing confidence intervals via data perturbation can be used as a system identification approach to estimate the parameter of linear regression models. Such a method is called Sign-perturbed Sums (SPS), which is a novel, distribution-free, finite-sample solution that relies only on mild statistical assumptions and provides non-asymptotic confidence set that contains the unknown system parameter with predetermined probability. The aim of this thesis was to contribute to the vast theory of Sign-perturbed Sums method and to analyze the possible applicability of SPS confidence regions in stochastic multi-armed bandit (MAB) problems.

We gave an overview of the construction and many known properties of SPS confidence regions (e.g. star convexity). In addition, we gave lower bounds for the probability of the events that the SPS confidence region is empty or becomes the whole space. The lower estimates showed that for small sample sizes and relatively high confidence level the probability of these events are not negligible. In a later part of the thesis, we also proved that these lower bounds are upper bounds as well in case of one-dimensional SPS (in case of empty regions, continuous noise is also required). Furthermore, we considered several modifications of the SPS method in general. We proved that the exact, user-chosen confidence level holds not only when sign-perturbation is applied but even if the perturbation variables are symmetric about zero (and they take the value 0 with probability 0).

In general, the construction of SPS confidence sets is computationally demanding, thus they are typically approximated. In the final parts of the thesis, we analyzed the one-dimensional SPS confidence regions, which are intervals. In that case, the confidence sets can be calculated in closed form. We presented an algorithm that determines these intervals. Furthermore, we introduced the concept of linear SPS, which provides semi-infinite intervals as confidence regions. Using linear SPS, we proposed a new Upper Confidence Bound (UCB) algorithm for symmetric MAB environments. At the end of the thesis, we empirically compared our bandit algorithm to the Asymptotically Optimal Upper Confidence Bound policy. After conducting numerical experiments using the Python programming language, we concluded that our linear SPS based UCB algorithm performs efficiently in terms of regret, especially in harder tasks, i.e. when the suboptimality gaps of the examined MAB are small.

2021