

ASYMPTOTIC NORMALITY OF ROBUST RISK MINIMIZERS

Stanislav Minsker*

University of Southern California

Abstract: This paper investigates the asymptotic properties of algorithms that can be viewed as robust analogues of the classical empirical risk minimization. These strategies are based on replacing the usual empirical average with a robust proxy of the mean, such as a variant of the median-of-means estimator. It is well known by now that the excess risk of resulting estimators often converges to zero at optimal rates under much weaker assumptions than those required by their classical counterparts. However, less is known about the asymptotic properties of the estimators themselves, for instance, whether robust analogues of the maximum likelihood estimators are asymptotically efficient. We take a step towards answering these questions and show that for a wide class of parametric problems, minimizers of the appropriately defined robust proxy of the risk converge to the minimizers of the true risk at the same rate, and often have the same asymptotic variance, as the estimators obtained by minimizing the usual empirical risk. Finally, we discuss the computational aspects of the problem and demonstrate the numerical performance of the methods under consideration in numerical experiments.

Key words and phrases: Adversarial contamination, asymptotic normality, median-of-means estimator, robust estimation.

1. Introduction

The concept of robustness addresses stability of statistical estimators under various forms of perturbations, such as the presence of corrupted/atypical observations (“outliers”) in the data. The questions related to robustness in the framework of statistical learning theory have seen a surge in interest, both from the theoretical and practical perspectives, and resulted in the development of novel algorithms. These new robust algorithms are characterized by the fact that they provably work under minimal assumptions on the underlying data-generating mechanism, often requiring the existence of moments of low order only. Majority of the existing works focused on the upper bounds for the risk of the estimators (such as the classification or prediction error) produced by the algorithms, while in this paper we are interested in the asymptotic properties of the estimators themselves. The asymptotic viewpoint allows one to gauge efficiency of the estimators and understand the magnitude of constants appearing

*Corresponding author. E-mail: minsker@usc.edu