Statistica Sinica **37** (2027), 1-23
doi:https://doi.org/10.5705/ss.202023.0329

EMPIRICAL RISK MINIMIZATION FOR LOSSES WITHOUT VARIANCE

Guanhua Fang*, Ping Li and Gennady Samorodnitsky

Fudan University, VecML Inc. and Cornell University

Abstract: This paper considers an empirical risk minimization problem under heavy-tailed settings, where data does not have finite variance, but only has $(1+\varepsilon)$-th moment with $\varepsilon \in (0,1)$. Instead of using an estimation procedure based on truncated observed data, we choose the optimizer by minimizing the risk value. Those risk values can be robustly estimated via using the remarkable Catoni's method. Thanks to the structure of Catoni-type influence functions, we are able to establish excess risk upper bounds via using generalized generic chaining methods. Moreover, we take computational issues into consideration. We especially theoretically investigate two types of optimization methods, robust gradient descent algorithm and empirical risk-based methods. With an extensive numerical study, we find that the optimizer based on empirical risks via Catoni-style estimation indeed shows better performance than other baselines. It indicates that estimation directly based on truncated data may lead to unsatisfactory results.

Key words and phrases: Catoni estimator, chaining method, excess risk, heavy tail, robust gradient descent.

1. Introduction

Modern data usually exhibit heavy-tail phenomena. For example, in the financial market (Bradley and Taqqu, 2003; Ahn, Kim and Ramaswami, 2012), the returns are usually not normally distributed. In telecommunications (Crovella and Taqqu, 1999; Glaz, Naus and Wallenstein, 2001), the data sources may sometimes experience a burst of extreme events. In network analysis, the distributions of indegrees, outdegrees, and sizes of connected components might be heavy-tailed (Meusel et al., 2014). In past decades, theoretical analysis of heavy-tailed data has increasingly become a hot topic in the machine learning field, including multi-armed bandits (Bubeck, Cesa-Bianchi and Lugosi, 2013), reinforcement learning (Zhuang and Sui, 2021), mean-estimation problems (Minsker, 2018; Lugosi and Mendelson, 2019), etc. Among those, the empirical risk minimization theory for heavy-tailed data has not been fully explored yet, especially in the situation when the variance does not exist. In this paper, we provide a corresponding theoretical framework standing on the remarkable estimator introduced by Catoni (2012).

*Corresponding author. E-mail: fanggh@fudan.edu.cn