



Robust estimation for nonrandomly distributed data

Shaomin Li¹ · Kangning Wang² · Yong Xu³

Received: 4 May 2021 / Revised: 8 July 2022 / Accepted: 16 August 2022 /

Published online: 12 October 2022

© The Institute of Statistical Mathematics, Tokyo 2022

Abstract

In recent years, many methodologies for distributed data have been developed. However, there are two problems. First, most of these methods require the data to be randomly and uniformly distributed across different machines. Second, the methods are mainly not robust. To solve these problems, we propose a distributed pilot modal regression estimator, which achieves robustness and can adapt when the data are stored nonrandomly. First, we collect a random pilot sample from different machines; then, we approximate the global MR objective function by a communication-efficient surrogate that can be efficiently evaluated by the pilot sample and the local gradients. The final estimator is obtained by minimizing the surrogate function in the master machine, while the other machines only need to calculate their gradients. Theoretical results show the new estimator is asymptotically efficient as the global MR estimator. Simulation studies illustrate the utility of the proposed approach.

Keywords Distributed data · Communication-efficient · Modal regression · Robustness

✉ Kangning Wang
wknsuda@126.com

✉ Yong Xu
xuyong0410@126.com

Shaomin Li
lsmjim@bnu.edu.cn

¹ Center for Statistics and Data Science, Beijing Normal University, No.18 Jinfeng Road, Zhuhai 519087, China

² School of Statistics, Shandong Technology and Business University, No.191 Binhai Middle Road, Yantai 264005, China

³ School of Business Administration, Shandong Technology and Business University, No.191 Binhai Middle Road, Yantai 264005, China