Steland, Ansgar; Chang, Yuan-Tsung
High-confidence nonparametric fixed-width uncertainty intervals and applications to projected high-dimensional data and common mean estimation. (English) Zbl 1464.62362

Summary: Nonparametric two-stage procedures to construct fixed-width confidence intervals are studied to quantify uncertainty. It is shown that the validity of the random central limit theorem (RCLT) accompanied by a consistent and asymptotically unbiased estimator of the asymptotic variance already guarantees consistency and first-order as well as second-order efficiency of the two-stage procedures. This holds under the common asymptotics where the length of the confidence interval tends toward 0 as well as under the novel proposed high-confidence asymptotics where the confidence level tends toward 1. The approach is motivated by and applicable to data analysis from distributed big data with nonnegligible costs of data queries. The following problems are discussed: Fixed-width intervals for the mean, for a projection when observing high-dimensional data, and for the common mean when using nonlinear common mean estimators under order constraints. The procedures are investigated by simulations and illustrated by a real data analysis.

MSC:
62L10 Sequential statistical analysis
62G07 Density estimation
62G15 Nonparametric tolerance and confidence regions
62R07 Statistical aspects of big data and data science
60G40 Stopping times; optimal stopping problems; gambling theory

Keywords:
big data; data science; high-dimensional data; jackknife; sequential analysis; sequential sampling

Full Text: DOI

References:
[12] Mukhopadhyay, N.; Silva, B. M., Sequential Methods and Their Applications (2009), Boca Raton, FL: Taylor and Francis, Boca Raton, FL · Zbl 1277.62024


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.