

In this paper we address the problem of performing statistical inference for large scale data sets i.e., Big Data. The volume and dimensionality of the data may be so high that it cannot be processed or stored in a single computing node. We propose a scalable, statistically robust and computationally efficient bootstrap method, compatible with distributed processing and storage systems. Bootstrap resamples are constructed with smaller number of distinct data points on multiple disjoint subsets of data, similarly to the bag of little bootstrap method (BLB) [A. Kleiner, A. Talwalkar, P. Sarkar, and M. I. Jordan, "A scalable bootstrap for massive data," *J. Roy. Statist. Soc.: Ser. B (Statist. Methodol.)*, vol. 76, no. 4, pp. 795-816, 2014]. The disjoint subsets are significantly smaller than the original full data set and they may be processed in different storage and computing units in parallel. Then significant savings in computation is achieved by avoiding the recomputation of the estimator for each bootstrap sample. Instead, a computationally efficient fixed point estimation equation is analytically solved via a smart approximation following the Fast and Robust Bootstrap method (FRB) [M. Salibián-Barrera, S. Van Aelst, and G. Willems, "Fast and robust bootstrap," *Statist. Methods Appl.*, vol. 17, no. 1, pp. 41-71, 2008]. Our proposed bootstrap method facilitates the use of highly robust statistical methods in analyzing large scale data sets. The favorable statistical properties of the method are established analytically. Numerical examples demonstrate scalability, low complexity and robust statistical performance of the method in analyzing large data sets.