Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on June 8, 2023

As one of the most fundamental problems in statistics, robust location estimation has many prominent solutions, such as the symmetric trimmed mean, symmetric Winsorized mean, Hodges–Lehmann estimator, Huber M-estimator, and median of means. Recent studies suggest that their biases concerning the mean can be quite different in asymmetric distributions, but the underlying mechanisms largely remain unclear. This study establishes two forms of orderliness within a wide range of semiparametric distributions. Further deductions explain why the Winsorized mean typically has smaller biases compared to the trimmed mean; two sequences of semiparametric robust mean estimators emerge. Building on the γ-U-orderliness, the superiority of the median Hodges–Lehmann mean is discussed.

semiparametric | mean-median-mode inequality | asymptotic | unimodal | Hodges–Lehmann estimator

Classifying Distributions by the Signs of Derivatives

Let $\mathcal{P}_{\mathbb{R}}$ denote the set of all continuous distributions over $\mathbb{R}$ and $\mathcal{P}_{\mathbb{X}}$ denote the set of all discrete distributions over a countable set $\mathbb{X}$. The primary focus of this article will be on the class of continuous distributions, $\mathcal{P}_{\mathbb{R}}$. However, it's worth noting that most discussions and results can be extended to encompass the discrete case, $\mathcal{P}_{\mathbb{X}}$, unless explicitly specified otherwise. Besides fully and smoothly parameterizing them by a Euclidean parameter or merely assuming regularity conditions, there exist additional methods for classifying distributions based on their characteristics, such as their skewness, peakedness, modality, and supported interval. In 1956, Stein initiated the study of estimating parameters in the presence of infinite-dimensional nuisance shape parameters (1), a contribution later explicitly recognized as initiating the field of semiparametric statistics.

Data Availability. Data for Figure **??** are given in SI Dataset S1. All codes have been deposited in GitHub.

ACKNOWLEDGMENTS. I sincerely acknowledge the insightful comments from the editor which considerably elevated the lucidity and merit of this paper.

1. CM Stein, Efficient nonparametric testing and estimation in *Proceedings of the third Berkeley symposium on mathematical statistics and probability*. Vol. 1, pp. 187–195 (1956).