Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on June 6, 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

Hodges–Lehmann inequality and γ-U-orderliness

The Hodges–Lehmann estimator stands out as a very unique robust location estimator due to its definition being substantially dissimilar from conventional symmetric weighted averages. In their landmark paper, *Estimates of location based on rank tests*, Hodges and Lehmann (1) proposed two methods to compute the H-L estimator: the Wilcoxon score R-estimator and the median of pairwise means, with time complexities of $O(nlog(n))$ and $O(n^2)$, respectively. The Wilcoxon score R-estimator is an estimator based on signed-rank test, or R-estimator, (1) and was later independently discovered by Sen (2, 3). However, the median of pairwise means is a generalized L-statistic and a trimmed U-statistic, as classified by Serfling in his novel conceptualized study in 1984 (4). Serfling further advanced the understanding by generalizing the H-L kernel as $hl_k = \frac{1}{k}\sum_{i=1}^{k} x_i$, where $k \in \mathbb{N}$ (4). Here, the weighted H-L kernel is defined as $whl_k = \frac{\sum_{i=1}^{k} x_i \mathbf{w}_i}{\sum_{i=1}^{k} \mathbf{w}_i}$.

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. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
2. PK Sen, On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. *Biometrics* pp. 532–552 (1963).
3. M Ghosh, MJ Schell, PK Sen, A conversation with pranab kumar sen. *Stat. Sci.* pp. 548–564 (2008).
4. RJ Serfling, Generalized l-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).