Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on June 7, 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

In 1823, Gauss (1) proved that for any unimodal distribution, $|m-\mu| \leq \sqrt{\frac{3}{4}}\omega$ and $\sigma \leq \omega \leq 2\sigma$, where μ is the population mean, m is the population median, ω is the root mean square deviation from the mode, and σ is the population standard deviation. This pioneering work revealed that, despite the potential bias of the median, the most fundamental robust mean estimate, its deviation remains bounded in units of a scale parameter under certain assumptions. Bernard, Kazzi, and Vanduffel (2020) (2) further derived asymptotic bias bounds of any quantile for unimodal distributions with finite second moments, by reducing this optimization problem to a parametric one, which can be solved analytically. They showed that m has the smallest maximum distance to μ among all symmetric quantile averages (SQA_ϵ). Daniell, in 1920, (3) analyzed a class of estimators, linear combinations of order statistics, and identified that the ϵ-symmetric trimmed mean (STM_ϵ) belongs to this class. Another popular choice, the ϵ-symmetric Winsorized mean (SWM_ϵ), named after Winsor and introduced by Tukey (4) and Dixon (5) in 1960, is also an L-estimator.

1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae.* (Henricus Dieterich), (1823).
2. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
3. P Daniell, Observations weighted according to order. *Am. J. Math.* **42**, 222–236 (1920).
4. JW Tukey, A survey of sampling from contaminated distributions in *Contributions to probability and statistics.* (Stanford University Press), pp. 448–485 (1960).
5. WJ Dixon, Simplified Estimation from Censored Normal Samples. *The Annals Math. Stat.* **31**, 385 – 391 (1960).