Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on May 25, 2023

As one of the most fundamental problem 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 maximum biases concerning the mean can be quite different in asymmetric distributions, but the underlying mechanisms and average performance remain largely unclear. This study establishes several forms of orderliness among quantile averages, similar to the mean-median-mode inequality, within a wide range of semiparametric distributions. From this, a sequence of advanced robust mean estimators emerges, which also explains why the Winsorized mean and median of means typically have smaller biases compared to the trimmed mean. Building on the 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, $|m - \mu| \leq \sqrt{\frac{3}{4}}\omega$, where μ is the population mean, m is the population median, and ω is the root mean square deviation from the mode. This pioneering work revealed that despite potential bias in robust mean estimates, the deviation remains bounded in units of a scale parameter under certain assumptions. Bernard, Rüschendorf, and Vanduffel (2015) further derived asymptotic bias bounds of any quantile for any distribution with a finite second moment (2). Subsequently, Bernard, Kazzi, and Vanduffel (2020) further refined these bounds for unimodal distributions for unimodal distributions, by reducing this optimization problem to a parametric one, which can be solved analytically (3). They showed that m has the smallest maximum distance to μ among all symmetric quantile averages (SQA_ϵ).

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

1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae.* (Henricus Dieterich), (1823).
2. C Bernard, L Rüschendorf, S Vanduffel, Value-at-risk bounds with variance constraints. *J. Risk Insur.* **84**, 923–959 (2017).
3. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).