Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee[a,1]

[a]Institute of Biomathematics, Macau SAR 999078, China

This manuscript was compiled on March 19, 2023

As arguably one of the most fundamental problems in statistics, robust location estimation has many prominent solutions, such as the trimmed mean, Winsorized mean, Hodges–Lehmann estimator, Huber M-estimator, and median of means. Recent research findings suggest that their biases with respect to mean can be quite different in asymmetric distributions, but the underlying mechanisms remain largely unclear. Here, similar to the mean-median-mode inequality, it is proven that in the context of nearly all common unimodal distributions, there exists an orderliness of symmetric quantile averages with different breakdown points. Further deductions explain why the Winsorized mean and median of means generally have smaller biases compared to the trimmed mean. Building on the U-orderliness, the supremacy of weighted 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 with a finite second moment, $|m - \mu| \leq \sqrt{\frac{3}{4}}\omega$, where μ is the population mean, m is the population median, and ω is the root mean square deviation from the mode, M. Bernard, Kazzi, and Vanduffel (2020) (2) derived bias bounds for the ϵ-symmetric quantile average (SQA_ϵ) for unimodal distributions, building on the works of Karlin and Novikoff (1963) and Li, Shao, Wang, and Yang (2018) (3, 4). They showed that the population median, m, has the smallest maximum distance to the population mean, μ, among all symmetric quantile averages. Daniell, in 1920, (5) analyzed a class of estimators, linear combinations of order statistics, and identified that ϵ-symmetric trimmed mean (TM_ϵ) belongs to this class. Another popular choice, the ϵ-symmetric Winsorized mean (WM_ϵ), which was named after Winsor and introduced by Tukey (6) and Dixon (7) in 1960, is also an L-statistic. Without assuming unimodality, Bieniek (2016) derived exact bias upper bounds of the Winsorized mean based on Danielak and Rychlik's work (2003) on the trimmed mean and confirmed that the former is smaller than the latter (8, 9). In 1964, Huber (10) introduced an M-estimator that can minimize the impact of outliers on parameter estimation by combining the squared error loss and absolute error loss in its objective function. While some L-statistics are also M-statistics (e.g., the median), M-statistics are typically very different. Minsker (2018) (11) and Mathieu (2022) (12) studied the concentration bounds of M-statistics. Comparing these bounds to those of the trimmed mean (13), derived by Oliveira and Orenstein in 2019, it is clear that the worse-case performance of parametric M-statistics is generally inferior. In 1963, Hodges and Lehmann (14) proposed a class of nonparametric location estimators based on rank tests and, from the Wilcoxon signed-rank statistic (15), deduced the median of pairwise means as a robust location estimator for a symmetric population. The concept of median of means ($\mathrm{MoM}_{k,b}$, k is the number of size in each block, b is the number of blocks) was implicit several times in Nemirovsky and Yudin (1983) (16), Jerrum, Valiant, and Vazirani (1986), (17) and Alon, Matias and Szegedy (1996) (18)'s works. Having good performance even for distributions with infinite second moments, the advantages of MoM have received increasing attention over the past decade (19–26). Devroye, Lerasle, Lugosi, and Oliveira (2016) showed that MoM nears the optimum of nonparametric mean estimation with regards to concentration bounds when the distribution has a heavy tail (24). Asymptotically, the Hodges-Lehmann (H-L) estimator is equivalent to $\mathrm{MoM}_{k=2,b=\frac{n}{k}}$, and it can be seen as the pairwise mean distribution is approximated by the bootstrap and sampling without replacement, respectively (for the asymptotic validity, the reader is referred to the foundational works of Efron (1979) (27), Bickel and Freedman (1981, 1984) (28, 29), and Helmers, Janssen, and Veraverbeke (1990) (30)). Laforgue, Clemencon, and Bertail (2019) proposed the median of randomized means (26), i.e., rather than partition, sampling without replacement is used to build data subsets, and showed that it has better non-asymptotic sub-Gaussian property compared to the MoM, i.e., the bootstrap approach should be preferred.

Here, the ϵ,b-stratified mean is defined as

$$\mathrm{SM}_{\epsilon,b,n} := \frac{b}{n}\left(\sum_{j=1}^{\frac{b-1}{2b\epsilon}} \sum_{i_j=\frac{(2bj-b-1)n\epsilon}{b-1}+1}^{\frac{(2bj-b+1)n\epsilon}{b-1}} X_{i_j}\right),$$

where $X_1 \leq \ldots \leq X_n$ denote the order statistics of a sample of n independent and identically distributed random variables $X_1, \ldots, X_n$, $\epsilon \bmod \frac{2}{b-1} = 0$, $\frac{1}{\epsilon} \geq 9$, $b \in \mathbb{N}$. $n \geq \frac{b-1}{2\epsilon}$. If the subscript n is omitted, only the asymptotic behavior is considered. If b is omitted, $b = 3$ is assumed. In situations

Significance Statement

In 1964, van Zwet introduced the convex transformation order for comparing the skewness of two distributions. This paradigm shift played a fundamental role in defining robust measures of distributions, from spread to kurtosis. Here, rather than the stochastic ordering between two distributions, the orderliness of quantile averages within a distribution is investigated. By classifying distributions through inequalities, a series of sophisticated robust mean estimators are deduced. Nearly all common nonparametric robust location estimators are found to be special cases thereof.

T.L. designed research, performed research, analyzed data, and wrote the paper.

The author declares no competing interest.

[1]To whom correspondence should be addressed. E-mail: tl@biomathematics.org

where $n \bmod \frac{b-1}{2\epsilon} \neq 0$, a possible solution is to generate several smaller samples that satisfy the equality by sampling without replacement, and then computing the mean of all estimations. This procedure can be viewed as sampling smaller samples from the population several times, thereby preserves the original distribution. The basic idea of the stratified mean is to distribute the random variables into $\frac{b-1}{2\epsilon}$ blocks according to their order, then further sequentially group these blocks into b strata and compute the mean of the middle stratum, which is the median of means of each stratum. Therefore, the stratified mean is a type of stratum mean in stratified sampling introduced by Neyman in 1934 (31). Although the principle is similar to that of the median of means, without the random shift, the result is different from $\mathrm{MoM}_{k=\frac{n}{b},b}$. The median of means and stratified mean are consistent mean estimators if their asymptotic breakdown points are zero. However, if $\epsilon = \frac{1}{9}$, the biases of the $\mathrm{SM}_{\frac{1}{9}}$ and the $\mathrm{WM}_{\frac{1}{9}}$ are almost indistinguishable in common asymmetric distributions (Figure ??, if no other subscripts, ϵ is omitted for simplicity), meaning that their robustness to departures from the symmetry assumption is practically similar. More importantly, the bounds confirm that the worst-case performance of Winsorized mean are better than those of the trimmed mean in terms of bias (8, 9), the complexity of bound analysis makes it difficult to achieve a complete and intuitive understanding. The aim of this paper is to define a series of semiparametric models using inequalities, reveal their elegant interrelations and connections to parametric models, and show that by exploiting these models, a set of sophisticated robust mean estimators can be deduced, which have strong robustness to departures from assumptions.

Quantile average and weighted average

ϵ-symmetric trimmed mean, ϵ-symmetric Winsorized mean, and ϵ-stratified mean are all L-statistics. More specifically, they are symmetric weighted averages, which is defined as

$$\mathrm{SWA}_{\epsilon,n} := \frac{\sum_{i=1}^{\frac{n}{2}} \frac{X_i + X_{n-i+1}}{2} w_i}{\sum_{i=1}^{\frac{n}{2}} w_i},$$

where w_is are the weights applied to the symmetric quantile averages according to the definition of the corresponding L-statistic. For example, for the ϵ-symmetric trimmed mean, $w_i = \begin{cases} 0, & i < n\epsilon \\ 1, & i \geq n\epsilon \end{cases}$. Mean and median are two special cases of symmetric trimmed mean ($\lim_{\epsilon \to 0} \mathrm{TM}_\epsilon$ and $\mathrm{TM}_{\frac{1}{2}}$, respectively). In 1974, Hogg investigated the asymmetric trimmed mean and found its advantages for some special applications (32). To extend to the asymmetric case, there are two possible definitions for the ϵ,γ-quantile average ($\mathrm{QA}(\epsilon, \gamma, n)$), i.e., $\frac{1}{2}(\hat{Q}_n(\gamma\epsilon) + \hat{Q}_n(1-\epsilon))$ and $\frac{1}{2}(\hat{Q}_n(\epsilon) + \hat{Q}_n(1-\gamma\epsilon))$, where $\gamma \geq 0$ and $0 \leq \epsilon \leq \frac{1}{1+\gamma}$, $\hat{Q}_n(p)$ is the empirical quantile function. For trimming from both sides, they are equivalent. For the sake of brevity, only the former is considered here, since many common asymmetric distributions are right skewed, the former allows trimming only from the right side by setting $\gamma = 0$. If γ is not specified, it is assumed to be 1. The symmetric quantile average is a special case of the quantile average when $\gamma = 1$.

Analogously, the weighted average can be defined as

$$\mathrm{WA}_{\epsilon,\gamma} := \frac{\int_{\epsilon_0=0}^{\frac{1}{1+\gamma}} \mathrm{QA}\left(\epsilon_0, \gamma\right) w_{\epsilon_0}}{\int_{\epsilon_0=0}^{\frac{1}{1+\gamma}} w_{\epsilon_0}}.$$

Converting this asymptotic definition to finite sample definition requires rounding the $n\epsilon_0$. For simplicity, only the asymptotic definition is considered here. For instance, the ϵ,γ-trimmed mean ($\mathrm{TM}_{\epsilon,\gamma}$) is a weighted average that trims the left side of the sample by $\gamma\epsilon$ and trims the right side of the sample by ϵ, where $w_{\epsilon_0} = \begin{cases} 0, & \epsilon_0 < \epsilon \\ 1, & \epsilon_0 \geq \epsilon \end{cases}$. Noted that a weighted average is an L-statistic, but an L-statistic might not be a weighted average, because in a weighted average, all quantiles are paired with the same γ.

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

ACKNOWLEDGMENTS. I gratefully acknowledge the valuable comments by the editor which substantially improved the clarity and quality of this paper.

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. S Karlin, A Novikoff, Generalized convex inequalities. *Pac. J. Math.* **13**, 1251–1279 (1963).
4. L Li, H Shao, R Wang, J Yang, Worst-case range value-at-risk with partial information. *SIAM J. on Financial Math.* **9**, 190–218 (2018).
5. P Daniell, Observations weighted according to order. *Am. J. Math.* **42**, 222–236 (1920).
6. JW Tukey, A survey of sampling from contaminated distributions in *Contributions to probability and statistics.* (Stanford University Press), pp. 448–485 (1960).
7. WJ Dixon, Simplified Estimation from Censored Normal Samples. *The Annals Math. Stat.* **31**, 385 – 391 (1960).
8. M Bieniek, Comparison of the bias of trimmed and winsorized means. *Commun. Stat. Methods* **45**, 6641–6650 (2016).
9. K Danielak, T Rychlik, Theory & methods: Exact bounds for the bias of trimmed means. *Aust. & New Zealand J. Stat.* **45**, 83–96 (2003).
10. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
11. S Minsker, Uniform bounds for robust mean estimators. *arXiv preprint arXiv:1812.03523* (2018).
12. T Mathieu, Concentration study of m-estimators using the influence function. *Electron. J. Stat.* **16**, 3695–3750 (2022).
13. RI Oliveira, P Orenstein, The sub-gaussian property of trimmed means estimators. *Unpubl. IMPA* (2019).
14. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
15. F Wilcoxon, Individual comparisons by ranking methods. *Biom. Bull.* **1**, 80–83 (1945).
16. AS Nemirovskij, DB Yudin, *Problem complexity and method efficiency in optimization.* (Wiley-Interscience), (1983).
17. MR Jerrum, LG Valiant, VV Vazirani, Random generation of combinatorial structures from a uniform distribution. *Theor. computer science* **43**, 169–188 (1986).
18. N Alon, Y Matias, M Szegedy, The space complexity of approximating the frequency moments in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing.* pp. 20–29 (1996).
19. PL Bühlmann, Bagging, subagging and bragging for improving some prediction algorithms in *Research report/Seminar für Statistik, Eidgenössische Technische Hochschule (ETH).* (Seminar für Statistik, Eidgenössische Technische Hochschule (ETH), Zürich), Vol. 113, (2003).
20. JY Audibert, O Catoni, Robust linear least squares regression. *The Annals Stat.* **39**, 2766–2794 (2011).
21. D Hsu, S Sabato, Heavy-tailed regression with a generalized median-of-means in *International Conference on Machine Learning.* (PMLR), pp. 37–45 (2014).
22. S Minsker, Geometric median and robust estimation in banach spaces. *Bernoulli* **21**, 2308–2335 (2015).
23. C Brownlees, E Joly, G Lugosi, Empirical risk minimization for heavy-tailed losses. *The Annals Stat.* **43**, 2507–2536 (2015).
24. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.* **44**, 2695–2725 (2016).
25. E Joly, G Lugosi, Robust estimation of u-statistics. *Stoch. Process. their Appl.* **126**, 3760–3773 (2016).
26. P Laforgue, S Clémençon, P Bertail, On medians of (randomized) pairwise means in *International Conference on Machine Learning.* (PMLR), pp. 1272–1281 (2019).
27. B Efron, Bootstrap methods: Another look at the jackknife. *The Annals Stat.* **7**, 1–26 (1979).
28. PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. *The annals statistics* **9**, 1196–1217 (1981).

29. PJ Bickel, DA Freedman, Asymptotic normality and the bootstrap in stratified sampling. *The annals statistics* **12**, 470–482 (1984).
30. R Helmers, P Janssen, N Veraverbeke, *Bootstrapping U-quantiles*. (CWI. Department of Operations Research, Statistics, and System Theory [BS]), (1990).
31. J Neyman, On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. Royal Stat. Soc.* **97**, 558–606 (1934).
32. RV Hogg, Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *J. Am. Stat. Assoc.* **69**, 909–923 (1974).