

Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on May 19, 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 have 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. In this article, similar to the mean-median-mode inequality, it is proven that within the context of nearly all common unimodal distributions, there is an orderliness of symmetric quantile averages with varying breakdown points. Further deductions explain 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 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 . This pioneering work revealed that despite potential bias with respect to the mean in robust estimates, the deviation remains bounded in unit 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 the population median, m , has the smallest maximum distance to the population mean, 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 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. Bieniek (2016) derived exact bias upper bounds of the Winsorized mean based on Danielak and Rychlik's work (2003) on the trimmed mean for any distribution with a finite second moment and confirmed that the former is smaller than the latter (6, 7). In 1963, Hodges and Lehmann (8) proposed a class of nonparametric location estimators based on rank tests and, from the Wilcoxon signed-rank statistic (9), deduced the median of pairwise means as a robust location estimator for a symmetric population. Both L -statistics and R -statistics achieve robustness essentially by removing a certain proportion of extreme values. In 1964, Huber (10) generalized maximum likelihood estimation to the minimization of the sum of a specific loss function, which measures the residuals between the data points and the model's parameters. Some L -estimators are also M -estimators, e.g., the sample mean is an M -estimator with a squared error loss function, while the sample median is

an M -estimator with an absolute error loss function (10). The Huber M -estimator is obtained by applying the Huber loss function that combines elements of both squared error and absolute error to achieve robustness against gross errors and high efficiency for contaminated Gaussian distributions (10). Sun, Zhou, and Fan (2020) examined the concentration bounds of Huber M -estimator (11). Mathieu (2022) (12) further derived the concentration bounds of M -estimators and demonstrated that, by selecting the tuning parameter which depends on the variance, Huber M -estimator can also be a sub-Gaussian estimator. The concept of median of means ($MoM_{k,b} = \frac{n}{k}$, k is the number of size in each block, b is the number of blocks) was implicitly introduced several times in Nemirovsky and Yudin (1983) (13), Jerrum, Valiant, and Vazirani (1986), (14) and Alon, Matias and Szegedy (1996) (15)'s works. Given its good performance even for distributions with infinite second moments, MoM has received increasing attention over the past decade (16–23). Devroye, Lerasle, Lugosi, and Oliveira (2016) showed that MoM nears the optimum of sub-Gaussian mean estimation with regards to concentration bounds when the distribution has a heavy tail (21). For a comparison of concentration bounds of trimmed mean, Huber M -estimator, median of means and other relevant estimators, readers are directed to Gobet, Lerasle, and Métivier's paper (2022) (24). Laforgue, Clemencon, and Bertail (2019) proposed the median of randomized means ($MoRM_{k,b}$) (23), wherein, rather than partitioning, an arbitrary number, b , of blocks are built independently from the sample, and showed that MoRM has a better non-asymptotic sub-Gaussian property compared to MoM. In fact, asymptotically, the Hodges–Lehmann (H-L) estimator is equivalent to $MoM_{k=2,b=\frac{n}{k}}$ and $MoRM_{k=2,b}$, and they can be seen as the pairwise mean distribution is approximated by the sampling without replacement and bootstrap, respectively. For the asymptotic validity, readers are referred to the foundational works of Efron (1979) (25), Bickel and

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 examining the stochastic ordering between two distributions, the orderliness of quantile averages within a distribution is investigated. By classifying distributions through the signs of derivatives, 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.

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

71 Freedman (1981, 1984) (26, 27), and Helmers, Janssen, and
 72 Veraverbeke (1990) (28).

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

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

73 where $X_1 \leq \dots \leq X_n$ denote the order statistics of a sample
 74 of n independent and identically distributed random variables
 75 X_1, \dots, X_n . $b \in \mathbb{N}$, $b \geq 3$. The definition was further refined to
 76 guarantee the continuity of the breakdown point by incorporat-
 77 ing an additional block in the center when $\lfloor \frac{b-1}{2b\epsilon} \rfloor \bmod 2 = 0$,
 78 or by adjusting the central block when $\lfloor \frac{b-1}{2b\epsilon} \rfloor \bmod 2 = 1$ (SI
 79 Text). If the subscript n is omitted, only the asymptotic
 80 behavior is considered. If b is omitted, $b = 3$ is assumed.
 81 $SM_{\epsilon, b=3}$ is equivalent to STM_{ϵ} , when $\epsilon > \frac{1}{6}$. The basic idea
 82 of the stratified mean, when $\frac{b-1}{2\epsilon} \in \mathbb{N}$, $b \bmod 2 = 1$, is to dis-
 83 tribute the data into $\frac{b-1}{2\epsilon}$ equal-sized non-overlapping blocks
 84 according to their order, then further sequentially group these
 85 blocks into b equal-sized strata and compute the mean of the
 86 middle stratum, which is the median of means of each stratum.
 87 In situations where $i \bmod 1 \neq 0$, a potential solution is to
 88 generate multiple smaller samples that satisfy the equality by
 89 sampling without replacement, and subsequently calculate the
 90 mean of all estimations. The details of determining the sample
 91 size and sampling times are provided in the SI Text. Although
 92 the principle resembles that of the median of means, without
 93 the random shift, the result is different from $MoM_{k=\frac{n}{b}, b}$. Ad-
 94 ditionally, the stratified mean differs from the mean of the
 95 sample obtained through stratified sampling methods, intro-
 96 duced by Neymean (1934) (29) or ranked set sampling (30),
 97 introduced by McIntyre in 1952, as these sampling methods
 98 aim to obtain more representative samples or improve the
 99 efficiency of sample estimates, but the sample mean based
 100 on them are not robust. When $b \bmod 2 = 1$, the stratified
 101 mean can be regarded as replacing the other equal-sized strata
 102 with the middle stratum, which, in principle, is analogous to
 103 the Winsorized mean that replaces extreme values with less
 104 extreme percentiles. Furthermore, while the bounds confirm
 105 that the Winsorized mean and median of means outperform
 106 the trimmed mean (6, 7, 21, 24) in worst-case performance,
 107 the complexity of bound analysis makes it difficult to achieve a
 108 complete and intuitive understanding of these results. Also, a
 109 clear explanation for the average performance of them remains
 110 elusive. The aim of this paper is to define a series of semi-
 111 parametric models using the signs of derivatives, reveal their
 112 elegant interrelations and connections to parametric models,
 113 and show that by exploiting these models, a set of sophisti-
 114 cated mean estimators can be deduced, which exhibit strong
 115 robustness to departures from assumptions.

116 Quantile average and weighted average

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

$$SWA_{\epsilon, n} := \frac{\sum_{i=1}^{\lceil \frac{n}{2} \rceil} \frac{X_i + X_{n-i+1}}{2} w_i}{\sum_{i=1}^{\frac{n}{2}} w_i},$$

where w_i s are the weights applied to the symmetric quantile
 averages according to the definition of the corresponding L -
 estimators. For example, for the ϵ -symmetric trimmed mean,
 $w_i = \begin{cases} 0, & i < n\epsilon \\ 1, & i \geq n\epsilon \end{cases}$, provided that $n\epsilon \in \mathbb{N}$. The mean and
 median are indeed two special cases of the symmetric trimmed
 mean.

To extend the symmetric quantile average to the asymmet-
 ric case, there are two possible definitions for the ϵ, γ -quantile
 average (QA(ϵ, γ, n)), i.e.,

$$\frac{1}{2}(\hat{Q}_n(\gamma\epsilon) + \hat{Q}_n(1-\epsilon)), \quad [1]$$

and

$$\frac{1}{2}(\hat{Q}_n(\epsilon) + \hat{Q}_n(1-\gamma\epsilon)), \quad [2]$$

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, [1] and [2] are equiva-
 lent. [1] is assumed in this article unless otherwise specified,
 since many common asymmetric distributions are right skewed,
 and [1] allows trimming only from the right side by setting
 $\gamma = 0$.

Analogously, the weighted average can be defined as

$$WA_{\epsilon, \gamma} := \frac{\int_{\epsilon_0=0}^{\frac{1}{1+\gamma}} QA(\epsilon_0, \gamma) w_{\epsilon_0}}{\int_{\epsilon_0=0}^{\frac{1}{1+\gamma}} w_{\epsilon_0}}.$$

For instance, the ϵ, γ -trimmed mean ($TM_{\epsilon, \gamma}$) is a weighted
 average with a left trim size of $\gamma\epsilon n$ and a right trim size of ϵn ,
 where $w_{\epsilon_0} = \begin{cases} 0, & \epsilon_0 < \epsilon \\ 1, & \epsilon_0 \geq \epsilon \end{cases}$.

Classifying distributions by the signs of derivatives

Let \mathcal{P}_k denote the set of all distributions over \mathbb{R} whose mo-
 ments, from the first to the k th, are all finite. Without loss
 of generality, for all the classes discussed in the following,
 they are assumed to be subclasses of the nonparametric class
 of distributions $\mathcal{P}_k^* := \{\text{All continuous distribution } P \in \mathcal{P}_k\}$.
 Besides fully and smoothly parameterizing by a Euclidean pa-
 rameter or just assuming regularity conditions, there are many
 ways to classify distributions. In 1956, Stein initiated the
 problem of estimating parameters in the presence of an infinite
 dimensional nuisance shape parameter (31). A notable exam-
 ple discussed in his groundbreaking work was the estimation
 of the center of symmetry for an unknown symmetric distribu-
 tion. In 1993, Bickel, Klaassen, Ritov, and Wellner published
 an influential semiparametrics textbook (32) which system-
 atically categorized many common models into three classes:
 parametric, nonparametric, and semiparametric. Yet, there is
 another old and commonly encountered class of distributions
 that receives little attention in semiparametric literature: the
 unimodal distribution. It is a very unique semiparametric
 model because its definition is based on the signs of deriva-
 tives, i.e., assuming P is continuous, $(f'(x) > 0 \text{ for } x \leq M) \wedge$
 $(f'(x) < 0 \text{ for } x \geq M)$. Let \mathcal{P}_U denote the set of all unimodal
 distributions. There was a widespread misbelief that the me-
 dian of an arbitrary unimodal distribution always lies between
 its mean and mode until Runnenburg (1978) and van Zwet
 (1979) (33, 34) endeavored to determine sufficient conditions

for the inequality to hold, thereby implying the possibility of its violation. The class of distributions that satisfy the mean-median-mode inequality constitutes a subclass of \mathcal{P}_U . By analogy, a right-skewed distribution is called γ -ordered, if and only if

$$\forall 0 \leq \epsilon_1 \leq \epsilon_2 \leq \frac{1}{1+\gamma}, \text{QA}_{\epsilon_1, \gamma} \geq \text{QA}_{\epsilon_2, \gamma}.$$

The necessary and sufficient condition below hints at the relation between the mean-median-mode inequality and the orderliness.

Theorem .1. Let P_Υ^k represent an arbitrary distribution in the set \mathcal{P}_Υ^k . P_Υ^k is γ -ordered if and only if the probability density function (pdf) satisfies the inequality $f(Q(\gamma\epsilon)) \geq f(Q(1-\epsilon))$ for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ or $f(Q(\gamma\epsilon)) \leq f(Q(1-\epsilon))$ for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$, where $\gamma \geq 0$.

Proof. Without loss of generality, consider the case of right-skewed continuous distribution. From the definition of γ -ordered distribution, it is deduced that $\frac{Q(\gamma\epsilon-\delta)+Q(1-\epsilon+\delta)}{2} \geq \frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \Leftrightarrow Q(\gamma\epsilon-\delta) - Q(\gamma\epsilon) \geq Q(1-\epsilon) - Q(1-\epsilon+\delta) \Leftrightarrow Q'(1-\epsilon) \geq Q'(\gamma\epsilon)$, where δ is an infinitesimal quantity. Observing that the quantile function is the inverse function of the cumulative distribution function (cdf), $Q'(1-\epsilon) \geq Q'(\gamma\epsilon) \Leftrightarrow F'(Q(\gamma\epsilon)) \geq F'(Q(1-\epsilon))$, thereby completing the proof, given that the derivative of cdf is pdf. \square

According to Theorem .1, if a probability distribution is right skewed and monotonic, it will always be γ -ordered, provided $\gamma \geq 0$. For a right skewed continuous unimodal distribution, if $Q(\gamma\epsilon) > M$, the inequality $f(Q(\gamma\epsilon)) \geq f(Q(1-\epsilon))$ holds. The principle is extendable to unimodal-like distributions. Suppose there is a right skewed continuous multimodal distribution following the mean- γ -median-first mode inequality with several smaller modes on the right side, with the first mode, M , having the greatest probability density, and the γ -median, $Q(\frac{\gamma}{1+\gamma})$, falling within the first dominant mode (i.e., if $x > Q(\frac{\gamma}{1+\gamma})$, $f(Q(\frac{\gamma}{1+\gamma})) \geq f(x)$), then if $Q(\gamma\epsilon) > M$, the inequality $f(Q(\gamma\epsilon)) \geq f(Q(1-\epsilon))$ also holds. In other words, while a distribution following the mean- γ -median-mode inequality may not be strictly γ -ordered, the inequality that defines γ -orderliness remains valid for most quantile averages. The mean- γ -median-mode inequality can also indicate possible bounds for γ in practice, e.g., for any distributions, when $\gamma \rightarrow \infty$, the γ -median will be greater than the mean and the mode, when $\gamma \rightarrow 0$, the γ -median will be smaller than the mean and the mode.

Consider the sign of the derivative of the quantile average with respect to the breakdown point, the above definition of γ -orderliness can also be expressed as

$$\forall 0 \leq \epsilon \leq \frac{1}{1+\gamma}, \frac{\partial \text{QA}_{\epsilon, \gamma}}{\partial \epsilon} \leq 0.$$

The left-skewed case can be obtained by reversing the inequality $\frac{\partial \text{QA}_{\epsilon, \gamma}}{\partial \epsilon} \leq 0$ to $\frac{\partial \text{QA}_{\epsilon, \gamma}}{\partial \epsilon} \geq 0$ and employing the second definition of QA, as given in [2]; for simplicity, it will be omitted in the following discussion. If $\gamma = 1$, the γ -ordered distribution is referred to as ordered.

Furthermore, many common right-skewed distributions are partial bounded, indicating a convex behavior of the QA function when $\epsilon \rightarrow 0$. By further assuming convexity, the

second γ -orderliness can be defined as follows for a right-skewed distribution,

$$\forall 0 \leq \epsilon \leq \frac{1}{1+\gamma}, \frac{\partial^2 \text{QA}_{\epsilon, \gamma}}{\partial \epsilon^2} \geq 0 \wedge \frac{\partial \text{QA}_{\epsilon, \gamma}}{\partial \epsilon} \leq 0.$$

Analogously, the ν th γ -orderliness of a right-skewed distribution can be defined as $(-1)^\nu \frac{\partial^\nu \text{QA}_{\epsilon, \gamma}}{\partial \epsilon^\nu} \geq 0 \wedge \dots \wedge -\frac{\partial \text{QA}_{\epsilon, \gamma}}{\partial \epsilon} \geq 0$. If $\gamma = 1$, the ν th γ -orderliness is referred as ν th orderliness. Let \mathcal{P}_O denote the set of all distributions that are ordered and let \mathcal{P}_{O_ν} and $\mathcal{P}_{\gamma O_\nu}$ denote the sets of all distributions that are ν th ordered and ν th γ -ordered, respectively. Five parametric distributions in \mathcal{P}_U are detailed as examples here: Weibull, gamma, Pareto, lognormal and generalized Gaussian. When the shape parameter of the Weibull distribution falls within a certain range, it can be shown that the Weibull distribution belongs to $\mathcal{P}_U \cap \mathcal{P}_O \cap \mathcal{P}_{O_2} \cap \mathcal{P}_{O_3}$ (SI Text). The parameters that exclude it from the set correspond to cases when it is near-symmetric, as shown in the SI Text. Nevertheless, calculating the derivatives of the QA function is often intricate and, at times, challenging. The following theorems can be used to quickly identify parametric distributions in \mathcal{P}_O , \mathcal{P}_{O_ν} , and $\mathcal{P}_{\gamma O_\nu}$.

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. 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).
6. M Bieniek, Comparison of the bias of trimmed and winsorized means. *Commun. Stat. Methods* **45**, 6641–6650 (2016).
7. K Danielak, T Rychlik, Theory & methods: Exact bounds for the bias of trimmed means. *Aust. & New Zealand J. Stat.* **45**, 83–96 (2003).
8. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
9. F Wilcoxon, Individual comparisons by ranking methods. *Biom. Bull.* **1**, 80–83 (1945).
10. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
11. Q Sun, WX Zhou, J Fan, Adaptive huber regression. *J. Am. Stat. Assoc.* **115**, 254–265 (2020).
12. T Mathieu, Concentration study of m-estimators using the influence function. *Electron. J. Stat.* **16**, 3695–3750 (2022).
13. AS Nemirovskij, DB Yudin, *Problem complexity and method efficiency in optimization*. (Wiley-Interscience), (1983).
14. MR Jerrum, LG Valiant, VV Vazirani, Random generation of combinatorial structures from a uniform distribution. *Theor. computer science* **43**, 169–188 (1986).
15. 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).
16. 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).
17. JY Audibert, O Catoni, Robust linear least squares regression. *The Annals Stat.* **39**, 2766–2794 (2011).
18. D Hsu, S Sabato, Heavy-tailed regression with a generalized median-of-means in *International Conference on Machine Learning*. (PMLR), pp. 37–45 (2014).
19. S Minsker, Geometric median and robust estimation in banach spaces. *Bernoulli* **21**, 2308–2335 (2015).
20. C Brownlees, E Joly, G Lugosi, Empirical risk minimization for heavy-tailed losses. *The Annals Stat.* **43**, 2507–2536 (2015).
21. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.* **44**, 2695–2725 (2016).
22. E Joly, G Lugosi, Robust estimation of u-statistics. *Stoch. Process. their Appl.* **126**, 3760–3773 (2016).
23. P Laforgue, S Cléménçon, P Bertail, On medians of (randomized) pairwise means in *International Conference on Machine Learning*. (PMLR), pp. 1272–1281 (2019).

- 248 24. E Gobet, M Lerasle, D Métivier, Mean estimation for Randomized Quasi Monte Carlo method.
249 working paper or preprint (2022).
- 250 25. B Efron, Bootstrap methods: Another look at the jackknife. *The Annals Stat.* **7**, 1–26 (1979).
- 251 26. PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. *The annals statistics* **9**,
252 1196–1217 (1981).
- 253 27. PJ Bickel, DA Freedman, Asymptotic normality and the bootstrap in stratified sampling. *The*
254 *annals statistics* **12**, 470–482 (1984).
- 255 28. R Helmers, P Janssen, N Veraverbeke, *Bootstrapping U-quantiles*. (CWI. Department of
256 Operations Research, Statistics, and System Theory [BS]), (1990).
- 257 29. J Neyman, On the two different aspects of the representative method: The method of stratified
258 sampling and the method of purposive selection. *J. Royal Stat. Soc.* **97**, 558–606 (1934).
- 259 30. G McIntyre, A method for unbiased selective sampling, using ranked sets. *Aust. journal*
260 *agricultural research* **3**, 385–390 (1952).
- 261 31. C Stein, , et al., Efficient nonparametric testing and estimation in *Proceedings of the third*
262 *Berkeley symposium on mathematical statistics and probability*. Vol. 1, pp. 187–195 (1956).
- 263 32. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparamet-*
264 *ric models*. (Springer) Vol. 4, (1993).
- 265 33. JT Runnenburg, Mean, median, mode. *Stat. Neerlandica* **32**, 73–79 (1978).
- 266 34. Wv Zwet, Mean, median, mode ii. *Stat. Neerlandica* **33**, 1–5 (1979).

DRAFT