Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on May 15, 2023

As 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 concerning the mean can be quite different in asymmetric distributions, but the underlying mechanisms remain largely unclear. In this article, 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 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 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 ($\mathrm{MoM}_{k,b=\frac{n}{k}}$, 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 relavent estimators, readers are directed to Gobet, Lerasle, and Métivier's paper (2022) (24). Laforgue, Clemencon, and Bertail (2019) proposed the median of randomized means ($\mathrm{MoRM}_{k,b}$) (23), wherein, rather than partitioning, an arbitrary number, b, of blocks are built independently from the sample, and showed that MoRM has better non-asymptotic sub-Gaussian property compared to MoM. In fact, asymptotically, the Hodges-Lehmann (H-L) estimator is equivalent to $\mathrm{MoM}_{k=2,b=\frac{n}{k}}$ and $\mathrm{MoRM}_{k=2,b}$, 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 Freedman (1981, 1984) (26, 27), and Helmers, Janssen, and Veraverbeke (1990)

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 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.

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

(28).

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

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

where $X_1 \leq ... \leq X_n$ denote the order statistics of a sample of n independent and identically distributed random variables $X_1, \ldots, X_n$. $b \in \mathbb{N}, b \geq 3$. The definition was further refined to guarantee the continuity of the breakdown point by incorporating an additional block in the center when $\lfloor \frac{b-1}{2b\epsilon} \rfloor \bmod 2 = 0$, or by adjusting the central block when $\lfloor \frac{b-1}{2b\epsilon} \rfloor \bmod 2 = 1$ (SI Text). If the subscript n is omitted, only the asymptotic behavior is considered. If b is omitted, $b = 3$ is assumed. $\mathrm{SM}_{\epsilon,b=3}$ is equal to STM_ϵ, when $\epsilon > \frac{1}{6}$. The basic idea of the stratified mean, when $\frac{b-1}{2\epsilon} \in \mathbb{N}$, $b \bmod 2 = 1$ is to distribute the data into $\frac{b-1}{2\epsilon}$ equal-sized non-overlapping blocks according to their order, then further sequentially group these blocks into b equal-sized strata and compute the mean of the middle stratum, which is the median of means of each stratum. In situations where $i \bmod 1 \neq 0$, a potential solution is to generate multiple smaller samples that satisfy the equality by sampling without replacement, and subsequently calculate the mean of all estimations, the details of determining the sample size and sampling times are included in the SI Text. Although the principle is similar to that of the median of means, without the random shift, the result is different from $\mathrm{MoM}_{k=\frac{n}{b},b}$. Additionally, the stratified mean differs from the mean of the sample obtained through stratified sampling methods, introduced by Neymean (1934) (29) or ranked set sampling (30), introduced by McIntyre in 1952, as these sampling methods are designed to obtain more representative samples or improve the efficiency of sample estimates, but the sample mean based on them are not robust. When $b \bmod 2 = 1$, the stratified mean can be regarded as replacing the other equal-sized strata with the middle stratum, which, in principle, is analogous to the Winsorized mean that replaces extreme values with less extreme percentiles. Furthermore, while the bounds confirm that the Winsorized mean and median of means outperform the trimmed mean (6, 7, 21, 24) in worst-case performance, the complexity of bound analysis makes it difficult to achieve a complete and intuitive understanding of these results. Also, a clear explanation for the average performance of them remains elusive. The aim of this paper is to define a series of semiparametric models using the signs of derivatives, 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

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

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

where w_is 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 asymmetric case, there are two possible definitions for the ϵ,γ-quantile average ($\mathrm{QA}(\epsilon, \gamma, n)$), i.e.,

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

and

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

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 equivalent. [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

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

For instance, the ϵ,γ-trimmed mean ($\mathrm{TM}_{\epsilon,\gamma}$) 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 moments, from the first to the kth, are all finite. Without loss of generality, all classes discussed in the following are subclasses of the nonparametric class of distributions $\mathcal{P}_\Upsilon^k \coloneqq \{\text{All continuous distribution } P \in \mathcal{P}_k\}$. Besides fully and smoothly parameterizing by a Euclidean parameter 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 example discussed in his groundbreaking work was the estimation of the center of symmetry for an unknown symmetric distribution. In 1993, Bickel, Klaassen, Ritov, and Wellner published an influential semiparametrics textbook (32) and systematically classified many common models into three classes: parametric, nonparametric, and semiparametric. However, 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 derivatives, 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. Five parametric distributions in $\mathcal{P}_U$ are detailed as examples here: Weibull, gamma, Pareto, lognormal and generalized Gaussian.

Inequalities related to weighted averages

The bias bound of the ϵ-symmetric trimmed mean is also monotonic for $\mathcal{P}_U \cap \mathcal{P}_2$, as proven in the SI Text using the formulae provided in Bernard et al.'s paper (2). So far, it appears clear that the bias of an estimator is closely related to its degree of robustness. For a right-skewed unimodal distribution, often, $\mu \geq \mathrm{STM}_\epsilon \geq m$. Then analogous to the γ-orderliness, the γ-trimming inequality is defined as $\forall 0 \leq \epsilon_1 \leq \epsilon_2 \leq \frac{1}{1+\gamma}, \mathrm{TM}_{\epsilon_1,\gamma} \geq \mathrm{TM}_{\epsilon_2,\gamma}$. A necessary and sufficient condition for the γ-trimming inequality is the monotonic decrease of the bias of the ϵ,γ-trimmed mean as a function of the breakdown point ϵ for a right skewed distribution, proven in the SI Text. γ-orderliness is a sufficient condition for the γ-trimming inequality, as proven in the SI Text, but it is not necessary.

Theorem .1. *For a right-skewed distribution following the γ-trimming inequality, asymptotically, the quantile average is always greater or equal to the corresponding trimmed mean with the same ϵ and γ, provided that $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ and $\gamma \geq 0$.*

Proof. Without loss of generality, assuming the distribution is continuous. According to the definition of the γ-trimming inequality: $\frac{1}{1-\epsilon-\gamma\epsilon+2\delta}\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du \geq \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$, if $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ and $\gamma \geq 0$, where δ is an infinitesimal positive quantity. Then, rewriting the inequality as $\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du - \frac{1-\epsilon-\gamma\epsilon+2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \geq 0 \Leftrightarrow \int_{1-\epsilon}^{1-\epsilon+\delta} Q(u)\,du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u)\,du - \frac{2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \geq 0$. Since $\delta \to 0^+$, $\frac{1}{2\delta}\left(\int_{1-\epsilon}^{1-\epsilon+\delta} Q(u)\,du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u)\,du\right) = \frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \geq \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$, the proof is complete. □

An analogous result can be obtained in the following theorem.

Theorem .2. *For a right-skewed continuous distribution following the γ-trimming inequality, asymptotically, the Winsorized mean is always greater or equal to the corresponding trimmed mean with the same ϵ and γ, provided that $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ and $0 \leq \gamma \leq 1$.*

Proof. According to Theorem .1, $\frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \geq \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \Leftrightarrow \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \geq \left(\frac{2\gamma\epsilon}{1-\epsilon-\gamma\epsilon}\right)\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$. Then, if $1 \geq \gamma \geq 0$, $\left(1-\frac{1}{1-\epsilon-\gamma\epsilon}\right)\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \geq 0 \Rightarrow \int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon) \geq \int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \geq \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$, the proof is complete. □

If assuming γ-orderliness, the result in Theorem .2 can be extended to the $\gamma > 1$ case, as proven in the SI Text. Replacing the trimmed mean in the γ-trimming inequality with Winsorized mean forms the definition of γ-Winsorization inequality. γ-orderliness also implies the γ-Winsorization inequality, if $0 \leq \gamma \leq 1$, as proven in the SI Text.

To construct weighted averages based on the γ-orderliness, let $\mathcal{B}_i = \int_{i\epsilon}^{(i+1)\epsilon} \mathrm{QA}(u,\gamma)\,du$, $ka = k\epsilon + c$, from the γ-orderliness, $-\frac{\partial \mathrm{QA}_{\epsilon,\gamma}}{\partial\epsilon} \geq 0 \Rightarrow \forall 0 \leq a \leq 2a \leq \frac{1}{1+\gamma}, -\frac{(\mathrm{QA}(2a,\gamma)-\mathrm{QA}(a,\gamma))}{a} \geq 0 \Rightarrow \mathcal{B}_i - \mathcal{B}_{i+1} \geq 0$. Let $\mathcal{B}_i = \mathcal{B}_0$, then, based on the γ-orderliness, ϵ,γ-block Winsorized mean, is defined here for comparison in the SI Dataset S1 as

$$\mathrm{BWM}_{\epsilon,\gamma,n} := \frac{1}{n}\left(\sum_{i=n\gamma\epsilon+1}^{(1-\epsilon)n} X_i + \sum_{i=n\gamma\epsilon+1}^{2n\gamma\epsilon+1} X_i + \sum_{i=(1-2\epsilon)n}^{(1-\epsilon)n} X_i\right),$$

which is double weighting the leftest and rightest blocks having the size $\gamma\epsilon n$ and ϵn. Since their sizes are different, the $0 \leq \gamma \leq 1$ is still necessary for the γ-block Winsorization inequality. If γ is omitted, $\gamma = 1$ is assumed. This terminology is the same for other weighted averages. The solutions for $i \bmod 1 \neq 0$ are the same as that in SM. From the second γ-orderliness, $\frac{\partial^2 \mathrm{QA}_{\epsilon,\gamma}}{\partial^2\epsilon} \geq 0 \Rightarrow \forall 0 \leq a \leq 2a \leq 3a \leq \frac{1}{1+\gamma}, \frac{1}{a}\left(\frac{(\mathrm{QA}(3a,\gamma)-\mathrm{QA}(2a,\gamma))}{a} - \frac{(\mathrm{QA}(2a,\gamma)-\mathrm{QA}(a,\gamma))}{a}\right) \geq 0 \Rightarrow \mathcal{B}_i - 2\mathcal{B}_{i+1} + \mathcal{B}_{i+2} \geq 0$. So, based on the second orderliness, SM_ϵ can be seen as assuming $\gamma = 1$, replacing the two blocks, $\mathcal{B}_i + \mathcal{B}_{i+2}$ with one block $2\mathcal{B}_{i+1}$. From the νth γ-orderliness, the recurrence relation of the derivatives naturally produces the alternating binomial coefficients,

$$(-1)^\nu \frac{\partial^\nu \mathrm{QA}_{\epsilon,\gamma}}{\partial\epsilon^\nu} \geq 0 \Rightarrow \forall 0 \leq a \leq \ldots \leq (\nu+1)a \leq \frac{1}{1+\gamma},$$

$$\frac{(-1)^\nu}{a}\left(\frac{\frac{\mathrm{QA}(\nu a+a,\gamma)^{\cdot^{\cdot^{\cdot}}}}{a} - \frac{^{\cdot^{\cdot^{\cdot}}}\mathrm{QA}(2a,\gamma)}{a}}{a} - \frac{\frac{\mathrm{QA}(\nu a,\gamma)^{\cdot^{\cdot^{\cdot}}}}{a} - \frac{^{\cdot^{\cdot^{\cdot}}}\mathrm{QA}(a,\gamma)}{a}}{a}\right)$$

$$\geq 0 \Leftrightarrow \frac{(-1)^\nu}{a^\nu}\left(\sum_{j=0}^{\nu}(-1)^j\binom{\nu}{j}\mathrm{QA}\left((\nu-j+1)a,\gamma\right)\right) \geq 0$$

$$\Rightarrow \sum_{j=0}^{\nu}(-1)^j\binom{\nu}{j}\mathcal{B}_{i+j} \geq 0.$$

Based on the νth orderliness, the ϵ-binomial mean is introduced as

$$\mathrm{BM}_{\nu,\epsilon,n} := \frac{1}{n}\left(\sum_{i=1}^{\frac{1}{2}\epsilon^{-1}(\nu+1)^{-1}}\sum_{j=0}^{\nu}\left(1-(-1)^j\binom{\nu}{j}\right)\mathfrak{B}_{i_j}\right),$$

where $\mathfrak{B}_{i_j} = \sum_{l=n\epsilon(j+(i-1)(\nu+1))+1}^{n\epsilon(j+(i-1)(\nu+1)+1)}\left(X_l + X_{n-l+1}\right)$. If ν is not indicated, it is default as $\nu = 3$. Since the alternating sum of binomial coefficients is zero, when $\nu \ll \epsilon^{-1}$, $\epsilon \to 0$, $\mathrm{BM} \to \mu$. If $\frac{1}{2}\epsilon^{-1}(\nu+1)^{-1} \in \mathbb{N}$, the asymmetry case is dividing the sample into ϵ^{-1} blocks in the same way as SM and then further weighting each block using binomial coefficients ($0 \leq \gamma \leq 1$ is needed). The solutions for the continuity of the breakdown point and $l \bmod 1 \neq 0$ are the same as that in SM and not repeated here. The equality $\mathrm{BM}_{\nu=1,\epsilon} = \mathrm{BWM}_\epsilon$ holds, and similarly, $\mathrm{BM}_{\nu=2,\epsilon} = \mathrm{SM}_{\epsilon,b=3}$, when $\gamma = 1$ and their respective ϵs are identical. Interestingly, the biases of the $\mathrm{SM}_{\epsilon=\frac{1}{9},b=3}$ and the $\mathrm{WM}_{\epsilon=\frac{1}{9}}$ are nearly indistinguishable in common asymmetric unimodal distributions such as Weibull, gamma, lognormal, and Pareto (SI Text), indicating that their robustness to departures from the symmetry assumption is practically similar. The reason is that the Winsorized mean is using two single quantiles to replace the trimmed parts, not two blocks. The following theorems explain this difference.

Theorem .3. *For a right-skewed γ-ordered continuous distribution, asymptotically, the Winsorized mean is always greater or equal to the corresponding block Winsorized mean with the same ϵ and γ, provided that $0 \leq \gamma \leq 1$.*

Proof. From the definitions of BWM and WM, removing the common part, $\sum_{i=n\gamma\epsilon+1}^{(1-\epsilon)n} X_i$, the statement requires $\lim_{n\to\infty}\left((n\gamma\epsilon)X_{n\gamma\epsilon+1}+(n\epsilon)X_{n-n\epsilon}\right) \geq \lim_{n\to\infty}\left(\sum_{i=n\gamma\epsilon+1}^{2n\gamma\epsilon} X_i+\sum_{i=n\epsilon}^{2n\epsilon-1} X_{n-i}\right)$. If $0\leq\gamma\leq 1$, every X_i can pair with a X_{n-i} to formed a quantile average, and the remaining X_{n-i}s are all smaller than $X_{n-n\epsilon}$, so the inequality is valid. ☐

If using single quantiles, based on the second γ-orderliness, the stratified quantile mean can be defined as

$$\mathrm{SQM}_{\epsilon,\gamma,n} := 4\epsilon\sum_{i=1}^{\frac{1}{4\epsilon}}\frac{1}{2}(\hat{Q}_n((2i-1)\gamma\epsilon)+\hat{Q}_n(1-(2i-1)\epsilon)),$$

$\mathrm{SQM}_{\epsilon=\frac{1}{4}}$ is the Tukey's midhinge (33). In fact, SQM is a subcase of SM when $\gamma=1$ and $b\to\infty$, so the solution for $\frac{1}{\epsilon} \bmod 4\neq 0$ is the same.

Theorem .4. *For a right-skewed second γ-ordered continuous distribution, asymptotically, $SQM_{\epsilon,\gamma}$ is always greater or equal to the corresponding $BM_{\nu=2,\epsilon,\gamma}$ with the same ϵ and γ, provided that $0\leq\gamma\leq 1$.*

Proof. For simplicity, suppose there are $\epsilon^{-1}\in\mathbb{N}$ blocks involving in the computation of both SQM and BM, The computation of $\mathrm{BM}_{\nu=2}$ involves alternating between weighting and non-weighting, let 0 means the block is assigned with a weight of zero, 1 means the block is assigned with a weighted of one, the sequence of denoting whether the block is weighted or not weighted is: $0,1,0,0,1,0,\ldots$. Let the sequence be denoted by $a_{\mathrm{BM}_{\nu=2}}(j)$, the formula for this sequence is $a_{\mathrm{BM}_{\nu=2}}(j)=\left\lfloor\frac{j \bmod 3}{2}\right\rfloor$. Similarly, the computation of SQM can be seen as placing quantiles (p) at the beginning of all blocks if $0<p<\frac{1}{1+\gamma}$, and at the end of all blocks if $p>\frac{1}{1+\gamma}$, the sequence of denoting whether the quantile in each block is weighted or not weighted is: $0,1,0,1,0,1,\ldots$. Let the sequence be denoted by $a_{\mathrm{SM}}(j)$, the formula for this sequence is $a_{\mathrm{SM}}(j)=j \bmod 2$. These sequences are also suitable if pairing all blocks and quantiles into block average, $\mathfrak{B}$, and quantile average, QA. There are two possible pairing of $a_{\mathrm{BM}_{\nu=2}}(j)$ and $a_{\mathrm{SM}}(j)$, one is $a_{\mathrm{BM}_{\nu=2}}(j)=a_{\mathrm{SM}}(j)=1$, another is $0,1,0$ in $a_{\mathrm{BM}_{\nu=2}}(j)$ pairing with $1,0,1$ in $a_{\mathrm{SM}}(j)$. By leveraging the same principle as Theorem .3 and the second γ-orderliness (replacing the two quantile averages with one quantile average in the middle), the desired result follows. ☐

The biases of $\mathrm{SQM}_{\epsilon=\frac{1}{8}}$, which is based on the second orderliness with a quantile approach, are also very similar to those of $\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{8}}$, which is based on the third orderliness with a block approach, in common asymmetric unimodal distributions (Figure **??**).

Hodges–Lehmann inequality and U-orderliness

The Hodges–Lehmann estimator is a very unique robust location estimator due to its definition being substantially dissimilar from conventional symmetric weighted averages. Hodges and Lehmann (8) in their landmark paper *Estimates of location based on rank tests* proposed two methods to compute the H-L estimator, Wilcoxon score R-estimator and median of pairwise means, whose time complexities are $O(nlog(n))$ and $O(n^2)$, respectively. The Wilcoxon score R-estimator is an estimator based on signed-rank test, or R-estimator (8), and was later independently discovered by Sen (34, 35). However, the median of pairwise means is a generalized L-statistic and a trimmed U-statistic (classified by Serfling in his novel conceptualized study in 1984) (36). He also further advanced the understanding by generalizing the H-L kernel, which is $hl_k=\frac{1}{k}\sum_{i=1}^{k}x_i$, where $k\in\mathbb{N}$ (36). By using the hl_k kernel and the weighted average, it is clear now that the Hodges-Lehmann estimator is also a weighted H-L mean, the definition of which is provided as follows,

$$\mathrm{WeHLM}_{k,\epsilon,\gamma,n} := \mathrm{WA}_{\epsilon_0,\gamma,n}\left(\left(hl_k\left(X_{N_1},\cdots,X_{N_k}\right)\right)_{N=1}^{\binom{n}{k}}\right),$$

where $\mathrm{WA}_{\epsilon_0,\gamma,n}(Y)$ denotes the ϵ_0,γ-weighted average with the sequence $\left(hl_k\left(X_{N_1},\cdots,X_{N_k}\right)\right)_{N=1}^{\binom{n}{k}}$ as an input. The asymptotic breakdown point of $\mathrm{WeHLM}_{k,\epsilon,\gamma}$ is $\epsilon=1-(1-\epsilon_0)^{\frac{1}{k}}$ (proven in another relavant paper). The bootstrap method can be used to ensure the continuity of k and therefore the breakdown point. Specifically, let the bootstrap size be denoted by b, then first sampling the original sample $(1-k+\lfloor k\rfloor)b$ times with the size of each sampling, $\lfloor k\rfloor$, and then subsequently sampling $(1-\lceil k\rceil+k)b$ times with the size of each sampling, $\lceil k\rceil$. The corresponding kernels are computed separately, and the pooled sequence is ultimately employed as an input for the WA. The $k=1$ case is the weighted average. Set the WA in WeHLM as TM_{ϵ_0}, it was named as trimmed H-L mean here (Figure **??**, $\epsilon_0=\frac{15}{64}$). $\mathrm{THLM}_{k=2}$ is close to the Wilcoxon's one-sample statistic investigated by Saleh in 1976 (37), which is first censoring the sample, and then computing the pairwise means. The hl_2 kernel distribution has a probability density function $f_{hl_2}(x)=\int_0^{2x}2f(t)f(2x-t)\,dt$ (a result after the transformation of variables), the support of the original distribution is assumed to be $[0,\infty)$ for simplicity. The expected value of the H-L estimator is the positive solution of $\int_0^{\text{H-L}}(f_{hl_2}(s))\,ds=\frac{1}{2}$. Due to the complexity of this equation, analytically proving the validity of the mean-H-L-median inequality for a distribution is hard. As an example, for the exponential distribution, $f_{hl_2}(x)=4\lambda^{-2}xe^{-2\lambda^{-1}x}$, $E[\text{H-L}]=\frac{-W_{-1}\left(-\frac{1}{2e}\right)-1}{2}\lambda\approx 0.839\lambda$, where W_{-1} is a branch of the Lambert W function.

By replacing the H-L kernel with the weighted H-L kernel, which is $whl_k=\frac{1}{k}\frac{\sum_{i=1}^{k}w_ix_i}{\sum_{i=1}^{k}w_i}$, the weighted L-statistic can be defined as follows

$$\mathrm{WL}_{k,\epsilon,\gamma,n} := \mathrm{WA}_{\epsilon_0,\gamma,n}\left(\left(whl_k\left(X_{N_1},\cdots,X_{N_k}\right)\right)_{N=1}^{\binom{n}{k}}\right).$$

The weighted H-L mean is a special case of the weighted L-statistic when all $w_i=1$. A complication arises when $w_i\neq 1$; in such cases, regardless of the choice of WA, the weighted L-statistic is not a consistent nonparametric mean estimator when $\epsilon_0=0$, so it is not detailed here. If replacing WA in WL with L-estimator, the resulting statistic is referred to as the LL-statistic.

Analogous to the trimming inequality, the Hodges-Lehmann inequality can be defined as $\forall k_2\geq k_1\geq 1, m\mathrm{HLM}_{k_2}\geq m\mathrm{HLM}_{k_1}$, where $m\mathrm{HLM}_k$ is setting the WA in WeHLM as median. Since $m\mathrm{HLM}_{k=1}=m$, $m\mathrm{HLM}_{k=2}=$H-L, $m\mathrm{HLM}_{k=\infty}=\mu$, if a distribution follows the H-L inequality, it also follows the mean-H-L-median inequality. Furthermore, the Hodges-Lehmann inequality is a special case of the γ-U-orderliness, i.e.,

$$(\forall k_2 \geq k_1 \geq 1, \mathrm{QHLM}_{k_2,\epsilon,\gamma} \geq \mathrm{QHLM}_{k_1,\epsilon,\gamma}) \vee$$
$$(\forall k_2 \geq k_1 \geq 1, \mathrm{QHLM}_{k_2,\epsilon,\gamma} \leq \mathrm{QHLM}_{k_1,\epsilon,\gamma}),$$

where QHLM_k is setting the WA in WeHLM as QA. The direction of the inequality depends on the relative magnitudes of $\mathrm{QA}_{\epsilon,\gamma}$ and μ, since $\mathrm{QHLM}_{k=1,\epsilon,\gamma} = \mathrm{QA}_{\epsilon,\gamma}$ and $\mathrm{QHLM}_{k=\infty,\epsilon,\gamma} = \mu$. U-orderliness is defined as setting $\gamma = 1$.

Theorem .5. *U-orderliness implies orderliness.*

Proof. The proof is demonstrated by establishing that a distribution exhibiting the U-orderliness property must be ordered. Suppose $n \to \infty$, $\frac{1}{2}\left(Q(0) + Q(1)\right) \geq \ldots \geq \frac{1}{2}\left(Q(\frac{i}{n}) + Q(\frac{n-i}{n})\right) \geq \ldots \geq Q(\frac{1}{2})$ is valid for a right-skewed ordered distribution. Let $\tilde{\epsilon} = \frac{i}{n}$, when $n \to \infty$, $\mathrm{SmQHLM}_{k=j,\tilde{\epsilon}\to 0,n} = \frac{1}{2j}\left(\sum_{i=0}^{j}\left(Q(\frac{i}{n}) + Q(\frac{n-i}{n})\right)\right)$, where SmQHLM is setting the WA in WeHLM as SQA. Since Theorem .1 implies that $\mu \leq \mathrm{SQA}_\epsilon$, when $\tilde{\epsilon} \to 0$, $\mathrm{SmQHLM}_{k_2,\tilde{\epsilon}} \leq \mathrm{SmQHLM}_{k_1,\tilde{\epsilon}}$ is equivalent to the orderliness. □

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 Clémençon, P Bertail, On medians of (randomized) pairwise means in *International Conference on Machine Learning.* (PMLR), pp. 1272–1281 (2019).
24. E Gobet, M Lerasle, D Métivier, Mean estimation for Randomized Quasi Monte Carlo method. working paper or preprint (2022).
25. B Efron, Bootstrap methods: Another look at the jackknife. *The Annals Stat.* **7**, 1–26 (1979).
26. PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. *The annals statistics* **9**, 1196–1217 (1981).
27. PJ Bickel, DA Freedman, Asymptotic normality and the bootstrap in stratified sampling. *The annals statistics* **12**, 470–482 (1984).
28. R Helmers, P Janssen, N Veraverbeke, *Bootstrapping U-quantiles.* (CWI. Department of Operations Research, Statistics, and System Theory [BS]), (1990).
29. 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).
30. G McIntyre, A method for unbiased selective sampling, using ranked sets. *Aust. journal agricultural research* **3**, 385–390 (1952).
31. C Stein, , et al., Efficient nonparametric testing and estimation in *Proceedings of the third Berkeley symposium on mathematical statistics and probability.* Vol. 1, pp. 187–195 (1956).
32. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparametric models.* (Springer) Vol. 4, (1993).
33. JW Tukey, , et al., *Exploratory data analysis.* (Reading, MA) Vol. 2, (1977).
34. PK Sen, On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. *Biometrics* pp. 532–552 (1963).
35. M Ghosh, MJ Schell, PK Sen, A conversation with pranab kumar sen. *Stat. Sci.* pp. 548–564 (2008).
36. RJ Serfling, Generalized l-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).
37. A Ehsanes Saleh, Hodges-lehmann estimate of the location parameter in censored samples. *Annals Inst. Stat. Math.* **28**, 235–247 (1976).