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 25, 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 based on loss function are typically very different. Minsker (2018) (11) and Mathieu (2022) (12) studied the concentration bounds of M-statistics. Comparing these bounds to that 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 ($\mathrm{MoM}_{k,b=\frac{n}{k}}$, 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). Laforgue, Clemencon, and Bertail (2019) proposed the median of randomized means ($\mathrm{MoRM}_{k=2,b}$) (26), i.e., rather than partition, an arbitrary number, b, of blocks are built independently from the sample, and showed that it has better non-asymptotic sub-Gaussian property compared to the MoM. In fact, asymptotically, the Hodges-Lehmann (H-L) estimator is equivalent to $\mathrm{MoRM}_{k=2,b}$ and $\mathrm{MoM}_{k=2,b=\frac{n}{k}}$, 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)).

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 $X_1 \leq \ldots \leq X_n$ 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 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

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 and median of means is better than that of the trimmed mean (8, 9, 13, 24, 26), the complexity of bound analysis makes it difficult to achieve a complete and intuitive understanding of these results. 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

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

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

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.,

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

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. For the sake of brevity, only [1] is considered here, since many common asymmetric distributions are right skewed, [1] 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

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

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.

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 (33). 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 (34) 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.

Consider the sign of the derivative of the quantile average with respect to the breakdown point, a right-skewed distribution is called γ-ordered, if and only if

$$\forall 0 \leq \epsilon \leq \frac{1}{1+\gamma}, \frac{d\mathrm{QA}}{d\epsilon} \leq 0.$$

It is reasonable, though not necessary, to further assume $\gamma \leq 1$ since the gross errors of a right-skewed distribution often come from the right side. The left-skewed case is obtained by reversing the inequality and using [2] as the definition of QA; for simplicity, it will be omitted in the following discussion. If $\gamma = 1$, the distribution is referred to as ordered. This nomenclature is used throughout the text. Let $\mathcal{P}_O$ denote the set of all ordered distributions. Pareto, lognormal, generalized Gaussian, and most cases of Weibull and gamma, are in $\mathcal{P}_U \cap \mathcal{P}_O$, as proven in the following discussion and SI Text. The only minor exceptions occur when the Weibull and gamma distribution are near-symmetric, as shown in the SI Text. In contrast to the mean-median-mode inequality, whose sufficient conditions are cumbersome, a necessary and sufficient condition of the γ-orderliness is the monotonic decreasing of the bias function of $\mathrm{QA}_{\epsilon,\gamma}$ with respect to ϵ for a right skewed distribution, as proven in the SI Text.

Furthermore, most common right-skewed distributions are partial bounded, indicating a convex decreasing behavior of the QA function when $\epsilon \to 0$. If assuming convexity further, the second γ-orderliness can be defined as the following for a right-skewed distribution plus the γ-orderliness,

$$\forall 0 \leq \epsilon \leq \frac{1}{1+\gamma}, \frac{d^2\mathrm{QA}}{d\epsilon^2} \geq 0 \wedge \frac{d\mathrm{QA}}{d\epsilon} \leq 0.$$

Analogously, the νth γ-orderliness can be defined as $(-1)^\nu \frac{d^\nu \mathrm{QA}}{d\epsilon^\nu} \geq 0 \wedge \ldots \wedge -\frac{d\mathrm{QA}}{d\epsilon} \geq 0$. The definition of νth orderliness is the same, just setting $\gamma = 1$. Common unimodal distributions are also second and third ordered, as shown in the SI Text. Let $\mathcal{P}_{O_\nu}$ and $\mathcal{P}_{\gamma O_\nu}$ denote the sets of all distributions that are νth ordered and νth γ-ordered. The following theorems can be used to quickly identify parametric distributions in $\mathcal{P}_{O_\nu}$ and $\mathcal{P}_{\gamma O_\nu}$.

Theorem .1. *Any symmetric distribution with a finite second moment is νth ordered.*

Proof. The assertion follows from the fact that for any symmetric distribution with a finite second moment, all symmetric quantile averages coincide. Therefore, the SQA function is always a horizontal line; the νth order derivative is zero. □

As a consequence of Theorem .1 and the fact that generalized Gaussian distribution is symmetric around the median, it is νth ordered.

Theorem .2. *Any continuous right skewed distribution whose Q satisfies $Q^{(\nu)}(p) \geq 0 \wedge \ldots Q^{(i)}(p) \geq 0 \ldots \wedge Q^{(2)}(p) \geq 0$, $i \bmod 2 = 0$, is νth γ-ordered, provided that $\gamma \leq 1$.*

Proof. Let $\mathrm{QA}(\epsilon) = \frac{1}{2}(Q(\gamma\epsilon) + Q(1-\epsilon))$, where $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ (also assumed in the following discussions for γ-ordered distributions), then $(-1)^j \frac{d^j \mathrm{QA}}{d\epsilon^j} = \frac{1}{2}((-\gamma)^j Q^j(\gamma\epsilon) + Q^j(1-\epsilon))$, $\nu \geq j \geq 1$. When $j \bmod 2 = 0$, $(-1)^j \frac{d^j \mathrm{QA}}{d\epsilon^j} \geq 0$, when $j \bmod 2 = 1$, the strict positivity is uncertain. If assuming $\gamma \leq 1$, $(-1)^j \frac{d^j \mathrm{QA}}{d\epsilon^j} \geq 0$, since $Q^{(j+1)}(\epsilon) \geq 0$. □

It is now trivial to prove that the Pareto distribution follows the νth γ-orderliness, provided that $\gamma \leq 1$, since the quantile function of the Pareto distribution is $Q(p) = x_m(1-p)^{-\frac{1}{\alpha}}$, $x_m > 0$, $\alpha > 0$, $Q^{(\nu)}(p) \geq 0$ according to the chain rule.

Theorem .3. *A right-skewed continuous distribution with a monotonic decreasing pdf is γ-ordered, if $\gamma \leq 1$.*

Proof. A monotonic decreasing pdf means $f'(x) = F^{(2)}(x) \leq 0$. Since $Q'(p) \geq 0$, let $x = Q(F(x))$, then by differentiating both sides of the equation twice, one can obtain $0 = Q^{(2)}(F(x))(F'(x))^2 + Q'(F(x))F^{(2)}(x) \Leftrightarrow Q^{(2)}(F(x)) = -\frac{Q'(F(x))F^{(2)}(x)}{(F'(x))^2} \geq 0$. The desired result is derived from Theorem .2. □

Theorem .3 gives a valuable insight into the relation between modality and γ-orderliness. The conventional definition states that a distribution with a monotonic pdf is still considered unimodal. However, within its supported interval, its mode number is zero. The number of modes and their magnitudes are closely related to the possibility of orderliness being valid, although counterexamples can always be constructed. A proof of γ-orderliness, if $\gamma \leq 1$, can be easily established for the gamma distributions when $\alpha \leq 1$ as the pdf of the gamma distribution is $f(x) = \frac{\lambda^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\lambda}}}{\Gamma(\alpha)}$, $x \geq 0$, $\lambda > 0$, $\alpha > 0$, which is a product of two monotonic decreasing functions under constraints. For $\alpha > 1$, the proof is challenging, numerical results show that the orderliness is valid until $\alpha > 140$ (SI Text), but it is instructive to consider that when $\alpha \to \infty$ the gamma distribution converges to a Gaussian distribution with mean $\mu = \alpha\lambda$ and variance $\sigma = \alpha\lambda^2$.

Theorem .4. *If transforming a symmetric unimodal random variable X with a function $\phi(x)$ such that $\frac{d^2\phi}{dx^2} \geq 0 \wedge \frac{d\phi}{dx} \geq 0$ over the interval supported, then the convex transformed distribution is ordered. If the quantile function of X satifies $Q^{(2)}(\epsilon) \leq 0$, the convex transformed distribution is second ordered.*

Proof. Let $\phi\mathrm{SQA}(\epsilon) = \frac{1}{2}(\phi(Q(\epsilon)) + \phi(Q(1-\epsilon)))$, then, $\frac{d\phi\mathrm{SQA}}{d\epsilon} = \frac{1}{2}(\phi'(Q(\epsilon))Q'(\epsilon) - \phi'(Q(1-\epsilon))Q'(1-\epsilon)) = \frac{1}{2}Q'(\epsilon)(\phi'(Q(\epsilon)) - \phi'(Q(1-\epsilon))) \leq 0$, since for a symmetric distribution, $m - Q(\epsilon) = Q(1-\epsilon) - m$, differentiating both sides, $-Q'(\epsilon) = -Q'(1-\epsilon)$, $Q'(\epsilon) \geq 0, \phi^{(2)} \geq 0$. Notably, differentiating twice, $Q^{(2)}(\epsilon) = -Q^{(2)}(1-\epsilon)$, $\frac{d^{(2)}\phi\mathrm{SQA}}{d\epsilon^{(2)}} = \frac{1}{2}\left(\left(\phi^{(2)}(Q(\epsilon)) + \phi^{(2)}(Q(1-\epsilon))\right)(Q'(\epsilon))^2\right) + \frac{1}{2}\left((\phi'(Q(\epsilon)) - \phi'(Q(1-\epsilon)))Q^{(2)}(\epsilon)\right)$. The sign of $\frac{d^{(2)}\phi\mathrm{SQA}}{d\epsilon^{(2)}}$ depends on $Q^{(2)}(\epsilon)$. □

The mean-median-mode inequality for distributions of the powers and roots of the variates of a given distribution was investigated by Henry Rietz in 1927 (35), but the most trivial solution is the exponential transformation since the derivatives are always positive. An application of Theorem .4 is that the lognormal distribution is ordered as it is exponentially transformed from the Gaussian distribution whose $Q^{(2)}(\epsilon) = -2\sqrt{2}\pi\sigma e^{2\mathrm{erfc}^{-1}(2\epsilon)^2}\mathrm{erfc}^{-1}(2\epsilon) \leq 0$ (so, it is also second ordered).

Theorem .4 also reveals a relation between convex transformation and orderliness, since ϕ is the non-decreasing convex function in van Zwet's trailblazing work *Convex transformations of random variables* (36). Consider a near-symmetric distribution S, such that SQA_ϵ as a function of ϵ fluctuates from 0 to $\frac{1}{2}$, with $\mu = m$. By definition, S is not ordered. Let s be the pdf of S. Applying the transformation $\phi(x)$ to S decreases $s(Q_S(\epsilon))$, and the decrease rate, due to the order, is much smaller for $s(Q_S(1-\epsilon))$. As a consequence, as the second derivative of $\phi(x)$ increases, eventually, after a point, $s(Q_S(\epsilon))$ becomes greater than $s(Q_S(1-\epsilon))$ even if it was not previously. Thus, the SQA_ϵ function becomes monotonically decreasing, and S becomes ordered. Accordingly, in a family of distributions that differ by a skewness-increasing transformation in van Zwet's sense, violations of orderliness typically occur only when the distribution is near-symmetric.

In 1895, Pearson proposed using the mean-median difference $\mu - m$ as a measure of skewness after standardization

(37). Bowley (1926) proposed a measure of skewness based on the SQA-median difference $\mathrm{SQA}_\epsilon - m$ (38). Groeneveld and Meeden (1984) (39) generalized these measures of skewness based on van Zwet's convex transformation (36) and investigated their properties. Similar to the orderliness, a distribution is called monotonically right skewed if and only if $\forall \epsilon_1 \le \epsilon_2 \le \frac{1}{2}, \mathrm{SQA}_{\epsilon_1} - m \ge \mathrm{SQA}_{\epsilon_2} - m$. Since m is a constant, the monotonic skewness is equivalent to the orderliness. The validity of robust measures of skewness based on the SQA-median difference is closely related to the orderliness of the distribution, because for a non-ordered distribution, the results from $\mathrm{SQA}_\epsilon - m$ with different breakdown points might be very different especially if the inequality $\mathrm{SQA}_\epsilon \ge m$ is not valid for some ϵ.

There was a widespread misbelief that the median of an arbitrary unimodal distribution always lies between its mean and mode until Runnenburg (1978) and van Zwet (1979) (40, 41) endeavored to determine sufficient conditions for the inequality to hold, thereby implying the possibility of its violation (counterexamples can be found in the papers by Dharmadhikari and Joag-Dev (1988), Basu and DasGupta (1997), and Abadir (2005)) (42–44). The class of distributions that satisfy the mean-median-mode inequality constitutes a subclass of $\mathcal{P}_U$. Analogously, the above definition of γ-orderliness can also be expressed as

$$\forall \epsilon_1 \le \epsilon_2 \le \frac{1}{1+\gamma}, \mathrm{QA}_{\epsilon_1,\gamma} \ge \mathrm{QA}_{\epsilon_2,\gamma}.$$

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

Theorem .5. *Let P_Υ^k denote an arbitrary distribution in the set $\mathcal{P}_\Upsilon^k$. $P_\Upsilon^k \in \mathcal{P}_{\gamma O}$ if and only if the pdf satisfies the inequality $f(Q(\gamma\epsilon)) \ge f(Q(1-\epsilon))$ for all $0 \le \epsilon \le \frac{1}{1+\gamma}$ or $f(Q(\gamma\epsilon)) \le f(Q(1-\epsilon))$ for all $0 \le \epsilon \le \frac{1}{1+\gamma}$, $Q(\epsilon)$ is the quantile function.*

Proof. Without loss of generality, just consider the right-skewed continuous case. From the definition of γ-ordered distribution, deducing $\frac{Q(\gamma\epsilon-\delta)+Q(1-\epsilon+\delta)}{2} \ge \frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \Leftrightarrow Q(\gamma\epsilon-\delta) - Q(\gamma\epsilon) \ge Q(1-\epsilon) - Q(1-\epsilon+\delta) \Leftrightarrow Q'(1-\epsilon) \ge Q'(\gamma\epsilon)$, where δ is an infinitesimal quantity. Since the quantile function is the inverse function of the cumulative distribution function (cdf), $Q'(1-\epsilon) \ge Q'(\gamma\epsilon) \Leftrightarrow F'(Q(\gamma\epsilon)) \ge F'(Q(1-\epsilon))$, the proof is complete by noticing that the derivative of cdf is the probability density function (pdf). □

According to Theorem .5, if a probability distribution is right skewed and monotonic, it will always be γ-ordered. While the same conclusion can be drawn from Theorem .3, Theorem .5 can be easily extended to the discrete case and the $\gamma > 1$ case. For a right skewed continuous unimodal distribution, if $Q(\gamma\epsilon) > M$, the inequality $f(Q(\gamma\epsilon)) \ge f(Q(1-\epsilon))$ holds. The principle can be extended to unimodal-like distributions as well. Suppose there is a right skewed continuous multimodal distribution following the mean-γ-median-first mode inequality with many small 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 $x > Q(\frac{\gamma}{1+\gamma})$, $f(Q(\frac{\gamma}{1+\gamma})) \ge f(x)$, then, if $Q(\gamma\epsilon) > M$, the inequality $f(Q(\gamma\epsilon)) \ge f(Q(1-\epsilon))$ 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 $\gamma \to \infty$, the γ-median will be greater than the mean and the mode, when $\gamma \to 0$, the γ-median will be smaller than the mean and the mode.

Remarkably, Bernard et al. (2020) (2) derived the bias bound of the symmetric quantile average for $\mathcal{P}_U$,

$$B_{\mathrm{SQAB}}(\epsilon) = \begin{cases} \frac{1}{2}\left(\sqrt{\frac{4}{9\epsilon} - 1} + \sqrt{\frac{3\epsilon}{4-3\epsilon}}\right) & \frac{1}{6} \ge \epsilon \ge 0 \\ \frac{1}{2}\left(\sqrt{\frac{1-\epsilon}{\epsilon+\frac{1}{3}}} + \sqrt{\frac{3\epsilon}{4-3\epsilon}}\right) & \frac{1}{2} \ge \epsilon > \frac{1}{6}. \end{cases}$$

The next theorem highlights its safeguarding role in defining estimators based on orderliness.

Theorem .6. *The above bias bound function, $B_{SQAB}(\epsilon)$, is monotonic decreasing over the interval $(0, \frac{1}{2})$.*

Proof. When $\frac{1}{6} \ge \epsilon \ge 0$, $B'_{\mathrm{SQAB}}(\epsilon) = \frac{1}{(4-3\epsilon)^2\sqrt{\frac{\epsilon}{12-9\epsilon}}} - \frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2}$. To prove $B'_{\mathrm{SQAB}} < 0$, it is equivalent to proving $(4-3\epsilon)^2\sqrt{\frac{\epsilon}{12-9\epsilon}} > 3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2$. Let $L(\epsilon) = (4-3\epsilon)^2\sqrt{\frac{\epsilon}{12-9\epsilon}}$, $R(\epsilon) = 3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2$, then $\frac{L(\epsilon)}{\epsilon^2} = \frac{(4-3\epsilon)^2}{\epsilon^2}\sqrt{\frac{\epsilon}{12-9\epsilon}} = \left(\frac{4}{\epsilon}-3\right)^2\sqrt{\frac{1}{\frac{12}{\epsilon}-9}}$, $\frac{R(\epsilon)}{\epsilon^2} = 3\sqrt{\frac{4}{\epsilon}-9}$. Assuming, $\frac{1}{\epsilon} \in \left(\frac{9}{4}, \infty\right)$, $\frac{L(\epsilon)}{\epsilon^2} > \frac{R(\epsilon)}{\epsilon^2} \Longleftrightarrow \left(\frac{4}{\epsilon}-3\right)^2\sqrt{\frac{1}{\frac{12}{\epsilon}-9}} > 3\sqrt{\frac{4}{\epsilon}-9} \Longleftrightarrow \left(\frac{4}{\epsilon}-3\right)^2 > 3\sqrt{\frac{4}{\epsilon}-9}\sqrt{\frac{12}{\epsilon}-9}$. Let $LmR\left(\frac{1}{\epsilon}\right) = \left(\frac{4}{\epsilon}-3\right)^4 - 9\left(\frac{4}{\epsilon}-9\right)\left(\frac{12}{\epsilon}-9\right)$, $\frac{dLmR(1/\epsilon)}{d(1/\epsilon)} = 32\left(32\left(\frac{1}{\epsilon}\right)^3 - 72\left(\frac{1}{\epsilon}\right)^2 + 27\frac{1}{\epsilon} + 27\right)$, $\frac{d^2LmR(1/\epsilon)}{d^2(1/\epsilon)} = 32\left(96\left(\frac{1}{\epsilon}\right)^2 - 144\left(\frac{1}{\epsilon}\right) + 27\right) > 0$, let $\frac{1}{\epsilon} = \frac{9}{4}$, $\frac{dLmR(1/\epsilon)}{d(1/\epsilon)} > 0$, therefore, $\frac{dLmR(1/\epsilon)}{d(1/\epsilon)} > 0$, for $\frac{1}{\epsilon} \in \left(\frac{9}{4}, \infty\right)$. Also, $LmR\left(\frac{9}{4}\right) > 0$, so, $LmR\left(\frac{1}{\epsilon}\right) > 0$ for $\epsilon \in \left(0, \frac{4}{9}\right)$. The first part is finished.

When $\frac{1}{2} \ge \epsilon > \frac{1}{6}$, $B'_{\mathrm{SQAB}}(\epsilon) = \frac{1}{(4-3\epsilon)^2\sqrt{\frac{\epsilon}{12-9\epsilon}}} - \frac{1}{(3\epsilon+1)^2\sqrt{\frac{1-\epsilon}{9\epsilon+3}}}$. To check whether $B'_{\mathrm{SQAB}}(\epsilon) < 0$, first using the two identities $\sqrt{\frac{1}{12-9\epsilon}} = \sqrt{\frac{1}{3(4-3\epsilon)}}$ and $\sqrt{\frac{1}{3+9\epsilon}} = \sqrt{\frac{1}{3(1+3\epsilon)}}$ to simplify the expression, and then the inequality becomes, $(4-3\epsilon)^{\frac{3}{2}}\sqrt{\epsilon} > (3\epsilon+1)^{\frac{3}{2}}\sqrt{1-\epsilon}\sqrt{\frac{1}{3}} \Longleftrightarrow (4-3\epsilon)^{\frac{3}{2}}\sqrt{\epsilon} > (3\epsilon+1)^{\frac{3}{2}}\sqrt{1-\epsilon}\sqrt{\frac{1}{3}} \Longleftrightarrow 3(4-3\epsilon)^3\epsilon > (3\epsilon+1)^3(1-\epsilon) \Longleftrightarrow -54\epsilon^4 + 324\epsilon^3 - 450\epsilon^2 + 184\epsilon - 1 > 0$. Since when $\epsilon < 1$, $-54\epsilon^4 + 54\epsilon^3 > 0$, just consider the condition that $270\epsilon^3 - 450\epsilon^2 + 184\epsilon - 1 > 0 \Longleftrightarrow \epsilon\left(270\epsilon^2 - 450\epsilon + 174\right) + 10\epsilon - 1 > 0$. Since $270\epsilon^2 - 450\epsilon + 174 > 0$ is valid for $\epsilon < \frac{1}{30}\left(25 - 3\sqrt{5}\right)$, so just need $10\epsilon - 1 > 0$, $10\epsilon > 1$, $\epsilon > \frac{1}{10}$. So, the inequality is valid for $\frac{1}{30}\left(25 - 3\sqrt{5}\right) \approx 0.610 > \epsilon > \frac{1}{10}$, within the range of $\frac{1}{2} \ge \epsilon > \frac{1}{6}$, therefore, $B'_{\mathrm{SQAB}} < 0$ for $\frac{1}{2} \ge \epsilon > \frac{1}{6}$. The first and second formula, when $\epsilon = \frac{1}{6}$, are all equal to $\frac{1}{2}\left(\sqrt{\frac{5}{3}} + \frac{1}{\sqrt{7}}\right)$. It follows that $B_{\mathrm{SQAB}}(\epsilon)$ is continuous over $\left(0, \frac{1}{2}\right)$. Hence, $B'_{\mathrm{SQAB}}(\epsilon) < 0$ is valid for $0 < \epsilon < \frac{1}{2}$, which leads to the assertion of this theorem. □

The proof is given in the SI Text. This monotonicity implies that the extent of any violations of the orderliness is bounded for a unimodal distribution, e.g., for a right-skewed unimodal distribution, if $\exists\epsilon_1 \le \epsilon_2 \le \epsilon_3 \le \frac{1}{2}$, $\mathrm{SQA}_{\epsilon_2} \ge \mathrm{SQA}_{\epsilon_3} \ge \mathrm{SQA}_{\epsilon_1}$, $\mathrm{SQA}_{\epsilon_2}$ will not be too far away from $\mathrm{SQA}_{\epsilon_1}$, since $\sup_{P\in\mathcal{P}_U\cap\mathcal{P}^2_\Upsilon}(\mathrm{SQA}_{\epsilon_1}) > \sup_{P\in\mathcal{P}_U\cap\mathcal{P}^2_\Upsilon}(\mathrm{SQA}_{\epsilon_2}) > \sup_{P\in\mathcal{P}_U\cap\mathcal{P}^2_\Upsilon}(\mathrm{SQA}_{\epsilon_3})$.

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, $\mu \ge \mathrm{TM}_\epsilon \ge m$. Then analogous to the γ-orderliness, the γ-trimming inequality is defined as $\forall\epsilon_1 \le \epsilon_2 \le \frac{1}{1+\gamma}, \mathrm{TM}_{\epsilon_1,\gamma} \ge \mathrm{TM}_{\epsilon_2,\gamma}$. A necessary and sufficient condition for the γ-trimming inequality is the monotonic decreasing 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, if $\gamma \ge 1$, as proven in the SI Text, but it is not necessary.

Theorem .7. *For a right-skewed continuous distribution following the γ-trimming inequality, the quantile average is always greater or equal to the corresponding trimmed mean with the same ϵ and γ.*

Proof. By deducing $\frac{1}{1-\epsilon-\gamma\epsilon+2\delta}\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du \ge \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \Leftrightarrow \int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u)\,du - \frac{1-\epsilon-\gamma\epsilon+2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \ge 0 \Leftrightarrow \int_{1-\epsilon}^{1-\epsilon+\delta} Q(u)\,du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u)\,du - \frac{2\delta}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \ge 0 \Leftrightarrow \frac{1}{2\delta}\left(\int_{1-\epsilon}^{1-\epsilon+\delta} Q(u)\,du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u)\,du\right) \ge \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$ and noticing that $\lim_{\delta\to 0}\frac{1}{2\delta}\left(\int_{1-\epsilon}^{1-\epsilon+\delta} Q(u)\,du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u)\,du\right) = \frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2}$, the proof is complete. □

A similar result can be obtained in the following theorem.

Theorem .8. *For a right-skewed continuous distribution following the γ-trimming inequality, the Winsorized mean is always greater or equal to the corresponding trimmed mean with the same ϵ and γ, provided that $\gamma \le 1$.*

Proof. Continue the above deduction, $\frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \ge \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du \Leftrightarrow \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \ge \left(\frac{2\gamma\epsilon}{1-\epsilon-\gamma\epsilon}\right)\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$. Then, if $\gamma \le 1$, $\left(1-\frac{1}{1-\epsilon-\gamma\epsilon}\right)\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \ge 0 \Rightarrow \int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon) \ge \int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du + \gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right) \ge \frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon} Q(u)\,du$, the proof is complete. □

Theorem .8 unequivocally establishes that, for a distribution that follows the γ-trimming inequality, the Winsorized mean is always less biased than the trimmed mean, as long as $\gamma \le 1$, except the γ-median, $Q\left(\frac{\gamma}{1+\gamma}\right)$, where they coincide. Additionally, the γ-orderliness is a sufficient condition of the γ-Winsorization inequality, if $\gamma \le 1$, as proven in the SI Text.

To construct weighted averages based on orderliness, let $\mathcal{B}_i = \int_{i\epsilon}^{(i+1)\epsilon}\mathrm{QA}(u)\,du$, $ka = k\epsilon + c$, from the γ-orderliness, $-\frac{d\mathrm{QA}}{d\epsilon} \ge 0 \Rightarrow \forall a \le 2a \le \frac{1}{1+\gamma}, -\frac{(\mathrm{QA}(2a)-\mathrm{QA}(a))}{a} \ge 0 \Rightarrow \mathcal{B}_i - \mathcal{B}_{i+1} \ge 0$. 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. $\gamma\epsilon n \in \mathbb{N}$. Since their sizes are different, the $\gamma \le 1$ is still necessary. If γ is omitted, it is assumed to be 1. From the second γ-orderliness $\frac{d^2\mathrm{QA}}{d^2\epsilon} \ge 0 \Rightarrow \forall a \le 2a \le 3a \le \frac{1}{1+\gamma}, \frac{1}{a}\left(\frac{(\mathrm{QA}(3a)-\mathrm{QA}(2a))}{a} - \frac{(\mathrm{QA}(2a)-\mathrm{QA}(a))}{a}\right) \ge 0 \Rightarrow \mathcal{B}_i - 2\mathcal{B}_{i+1} + \mathcal{B}_{i+2} \ge 0$. 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,

$$(-1)^\nu\frac{d^\nu\mathrm{QA}}{d\epsilon^\nu} \ge 0 \Rightarrow \forall a \le \ldots \le (\nu+1)a \le \frac{1}{2},$$

$$\frac{(-1)^\nu}{a}\left(\frac{\frac{\mathrm{QA}(\nu a+a)\ddots}{a} - \frac{\iddots\mathrm{QA}(2a)}{a}}{a} - \frac{\frac{\mathrm{QA}(\nu a)\ddots}{a} - \frac{\iddots\mathrm{QA}(a)}{a}}{a}\right)$$

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

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

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 $\mathfrak{B}_{i_j} = \sum_{l=n\epsilon(j+(i-1)(\nu+1))+1}^{n\epsilon(j+(i-1)(\nu+1)+1)}\left(X_l + X_{n-l+1}\right)$, $\frac{1}{2}\epsilon^{-1}(\nu+1)^{-1} \in \mathbb{N}$. If ν is not indicated, it is default as $\nu = 3$. Since the alternating sum of binomial coefficients is zero, when $\nu \ll \epsilon^{-1}$, $\epsilon \to 0$, $\mathrm{BM} \to \mu$. The solution for $n \bmod \epsilon^{-1} \neq 0$ is the same as that in the stratified mean. The asymmetry case is dividing the sample into ϵ^{-1} blocks in the same way as BWM and then further weighting each block using binomial coefficients ($\gamma \le 1$ is needed). The equality $\mathrm{BM}_{\nu=1,\epsilon} = \mathrm{BWM}_\epsilon$ holds, and similarly, $\mathrm{BM}_{\nu=2,\epsilon} = \mathrm{SM}_{\epsilon,b=3}$, when their respective ϵs are identical. The reason that $\mathrm{SM}_{\frac{1}{9}}$ has similar biases as $\mathrm{WM}_{\frac{1}{9}}$ is that the Winsorized mean is using two single quantiles to replace the trimmed parts, not two blocks. The following theorem explains this difference.

Theorem .9. *For a right-skewed γ-ordered continuous distribution, the Winsorized mean is always greater or equal to the corresponding block Winsorized mean with the same ϵ and γ, provided that $\gamma \le 1$.*

Proof. From the definitions of BWM and WM, removing the trimmed mean part, 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 $\gamma \leq 1$, 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

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

where $\frac{1}{\epsilon}$ mod $4=0$. $\mathrm{SQM}_{\frac{1}{4}}$ is the Tukey's midhinge (45).

Theorem .10. *For a right-skewed γ-ordered continuous distribution, $SQM_{\epsilon,\gamma}$ is always greater or equal to the corresponding $BM_{\nu=2,\epsilon,\gamma}$ with the same ϵ and γ, provided that $\gamma \leq 1$.*

Proof. When $\nu = 2$, the computation of BM involves alternating between weighting and non-weighting, i.e., the first block is assigned with a weight of zero, the second block is assigned with a non-zero weight, the third block is assigned with a zero weight, and so on. Therefore, every non-zero weighted block in $\mathrm{BM}_{\nu=2,\epsilon,\gamma}$ can be paired with a quantile in $\mathrm{SQM}_{\epsilon,\gamma}$ with the quantile positioned at the beginning of the block if $\epsilon < \frac{1}{1+\gamma}$, and at the end of the block if $\epsilon > \frac{1}{1+\gamma}$. By leveraging the same principle as Theorem .9, the desired result follows. □

The biases of $\mathrm{SQM}_{\frac{1}{8}}$, which is based on the second orderliness with single quantiles approach, are also very close to those of $\mathrm{BM}_{\frac{1}{8}}$ (Figure ??), which is based on the third orderliness with block approach.

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 (14) in their landmark paper *Estimates of location based on rank tests* proposed two methods to compute the H-L estimator, Wilcoxon score and median of pairwise means, whose time complexities are $O(nlog(n))$ and $O(n^2)$, respectively. The Wilcoxon score is an estimator based on signed-rank test, or R-statistic (14), and was later independently discovered by Sen (46, 47). However, the median of pairwise means is a generalized L-statistic (classified by Serfling in 1984) (48) and a trimmed U-statistic (classified by Janssen, Serfling, and Veraverbeke in 1987) (49). By modifying the hl_k kernel pointed by Janssen, Serfling and Veraverbeke in 1987 (49) and weighted average generalized here, it is clear now that the H-L estimator is a weighted H-L mean, the definition of which is provided as follows,

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

where $k \geq 1$, $\mathrm{WA}_{\epsilon_0,\gamma,n}(Y)$ 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. To ensure asymptotic continuity, $hl_k = \frac{1}{K}\sum_{i=1}^{K}x_i$, where $P(K=\lfloor k\rfloor)=1-k+\lfloor k\rfloor$, $P(K=\lceil k\rceil)=1-\lceil k\rceil+k$. The asymptotic breakdown point of $\mathrm{WHLM}_{k,\epsilon,\gamma}$ is $\epsilon = 1-(1-\epsilon_0)^{\frac{1}{k}}$ (proven in another relavant paper). The $k=1$ case is the weighted average. Set the WA 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 (50), which is first censoring the sample, and then computing the pairwise means. The hl_2 kernel distribution has a probability density function $f_{hl_2}(x) = \int_0^{2x}2f\left(t\right)f\left(2x-t\right)dt$ (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 $\int_0^{\text{H-L}}\left(f_{hl_2}(s)\right)ds = \frac{1}{2}$. 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 is the Lambert W function.

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 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 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 .11. *U-orderliness implies orderliness.*

Proof. Suppose n mod $2=0$, $n\to\infty$, $\frac{1}{2}\left(X_1+X_n\right) \geq \ldots \geq \frac{1}{2}\left(X_i+X_{n-i+1}\right) \geq \ldots \geq \frac{1}{2}\left(X_{\frac{n}{2}}+X_{\frac{n}{2}+1}\right)$ is valid for a sample from a right-skewed ordered distribution. Let $\tilde{\epsilon} = \frac{i}{n}$, when $\tilde{\epsilon}\to 0$, $\mathrm{SQHLM}_{k_2,\tilde{\epsilon}} \leq \mathrm{SQHLM}_{k_1,\tilde{\epsilon}}$ is equivalent to the orderliness, since $\mathrm{SQHLM}_{k=j,\tilde{\epsilon}\to 0,n} = \frac{1}{2j}\left(\sum_{i=1}^{j}(X_i+X_{n-i+1})\right)$ and Theorem .7 implies that $\mu \leq \mathrm{SQA}_{\epsilon}$. □

Be aware that the U-orderliness itself does not assume any orderliness within the hl_k kernel distribution. The $hl_{k=n-1}$ kernel distribution has n elements, and their order is the same as the original distribution, so it is ordered if and only if the original distribution is ordered. If assuming symmetry, the result is trivial since the k-fold convolutions of a symmetric distribution is also symmetric (proved by Laha in 1961)(51). However, proving other cases is challenging. For example, $f'_{hl_2}(x) = 4f\left(2x\right)f\left(0\right)+\int_0^{2x}4f\left(t\right)f'\left(2x-t\right)dt$, the strict negative of $f'_{hl_2}(x)$ is not guaranteed if just assuming $f'(x) < 0$, so, even if the original distribution is monotonic, the hl_2 kernel distribution might be non-monotonic. Also, unlike the pairwise difference distribution, if the original distribution is unimodal, the pairwise mean distribution might be non-unimodal, as demonstrated by a counterexample given by Chung in 1953 and mentioned by Hodges and Lehmann in 1954 (52, 53). If all hl_k kernel distributions, $k \geq 1$, are νth ordered, then the distribution is νth U-ordered. From that, the binomial H-L mean (set the WA as BM) is the bias-optimum weighted H-L mean, if

k is a constant (Figure ??), although its maximum breakdown point is ≈ 0.065 if $\nu = 3$. A comparison of the biases of $\mathrm{BM}_{\frac{1}{8}}$, $\mathrm{SQM}_{\frac{1}{8}}$, $\mathrm{THLM}_{k=2,\epsilon=\frac{1}{8}}$, $\mathrm{MHHLM}_{k=\frac{2\ln(2)-\ln(3)}{3\ln(2)-\ln(7)},\epsilon=\frac{1}{8}}$ (midhinge H-L mean), $m\mathrm{HLM}_{k=\frac{\ln(2)}{3\ln(2)-\ln(7)},\epsilon=\frac{1}{8}}$, $\mathrm{THLM}_{k=5,\epsilon=\frac{1}{8}}$, and $\mathrm{WHLM}_{k=5,\epsilon=\frac{1}{8}}$ is appropriate (Figure ??), given their comparable breakdown points, with $m\mathrm{HLM}_{k=5}$ exhibiting the third smallest biases. This result and Theorem .11 align with Devroye et al. (2016)'s seminal work that MoM is nearly optimal with regards to concentration bounds for heavy-tailed distributions (24), since when k is much smaller than n, the difference between sampling with replacement and without replacement is negligible, $m\mathrm{HLM}_{k,n}$ is asymptotically equivalent to $\mathrm{MoM}_{k,b=\frac{n}{k}}$ if assuming k is a constant. Hence, $\mathrm{MoM}_{k,b=\frac{n}{k}}$ is also based on U-orderliness.

In 1958, Richtmyer introduced the concept of quasi-Monte Carlo simulation that utilizes low-discrepancy sequences, resulting in a significant reduction in computational expenses for large sample simulation (54). Among various numerical sets, Sobol sequences are often favored in quasi-Monte Carlo methods (55). Building upon this principle, in 1991, Do and Hall extended it to bootstrap and found that the quasi-random approach resulted in lower variance compared to other bootstrap Monte Carlo procedures (56). By using a deterministic approach, the variance of $m\mathrm{HLM}_{k,n}$ is much lower than that of $\mathrm{MoM}_{k,b=\frac{n}{k}}$ (SI Dataset S1), when k is small. This highlights the superiority of the median Hodges-Lehmann mean over the median of means, as it not only can provide an accurate estimate for moderate sample sizes, but also allows the use of quasi-bootstrap, where the bootstrap size can be adjusted as needed.

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).
33. 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).
34. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparametric models.* (Springer) Vol. 4, (1993).
35. H Rietz, On certain properties of frequency distributions of the powers and roots of the variates of a given distribution. *Proc. Natl. Acad. Sci.* **13**, 817–820 (1927).
36. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
37. K Pearson, X. contributions to the mathematical theory of evolution.—ii. skew variation in homogeneous material. *Philos. Transactions Royal Soc. London.(A.)* **186**, 343–414 (1895).
38. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
39. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).
40. JT Runnenburg, Mean, median, mode. *Stat. Neerlandica* **32**, 73–79 (1978).
41. Wv Zwet, Mean, median, mode ii. *Stat. Neerlandica* **33**, 1–5 (1979).
42. S Basu, A DasGupta, The mean, median, and mode of unimodal distributions: a characterization. *Theory Probab. & Its Appl.* **41**, 210–223 (1997).
43. S Dharmadhikari, K Joag-Dev, *Unimodality, convexity, and applications.* (Elsevier), (1988).
44. KM Abadir, The mean-median-mode inequality: Counterexamples. *Econom. Theory* **21**, 477–482 (2005).
45. JW Tukey, , et al., *Exploratory data analysis.* (Reading, MA) Vol. 2, (1977).
46. PK Sen, On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. *Biometrics* pp. 532–552 (1963).
47. M Ghosh, MJ Schell, PK Sen, A conversation with pranab kumar sen. *Stat. Sci.* pp. 548–564 (2008).
48. RJ Serfling, Generalized l-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).
49. P Janssen, R Serfling, N Veraverbeke, Asymptotic normality of u-statistics based on trimmed samples. *J. statistical planning inference* **16**, 63–74 (1987).
50. A Ehsanes Saleh, Hodges-lehmann estimate of the location parameter in censored samples. *Annals Inst. Stat. Math.* **28**, 235–247 (1976).
51. RG Laha, On a class of unimodal distributions. *Proc. Am. Math. Soc.* **12**, 181–184 (1961).
52. J Hodges, E Lehmann, Matching in paired comparisons. *The Annals Math. Stat.* **25**, 787–791 (1954).
53. K Chung, Sur les lois de probabilité unimodales. *COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES* **236**, 583–584 (1953).
54. RD Richtmyer, A non-random sampling method, based on congruences, for" monte carlo" problems, (New York Univ., New York. Atomic Energy Commission Computing and Applied . . .), Technical report (1958).
55. IM Sobol', On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **7**, 784–802 (1967).
56. KA Do, P Hall, Quasi-random resampling for the bootstrap. *Stat. Comput.* **1**, 13–22 (1991).