

Semiparametric robust mean estimation 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 7, 2023

As arguably the most fundamental problem in statistics, nonparametric robust location estimation has many prominent solutions, such as the trimmed mean, Winsorized mean, Hodges–Lehmann estimator, and median of means. Recent research suggests that their biases with respect to mean can be quite different in asymmetric distributions. 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, 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 $_{\epsilon}$) 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 m has the smallest maximum distance to the 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 $_{\epsilon}$) belongs to this class. Another popular choice, the ϵ -symmetric Winsorized mean (WM $_{\epsilon}$), 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 1963, Hodges and Lehmann (10) proposed a class of nonparametric location estimators based on rank tests and, from the Wilcoxon signed-rank statistic (11), deduced the median of pairwise means as a robust location estimator for a symmetric population. The concept of median of means (MoM $_{k,b}$, k is the number of size in each block, b is the number of blocks) was implicit several times in Nemirovsky and Yudin (1983) (12), Jerrum, Valiant, and Vazirani (1986), (13) and Alon, Matias and Szegedy (1996) (14)'s works. Having good performance even for distributions with infinite second moments, the advantages of MoM have received increasing attention over the past decade (15–22). 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 (20). In fact, asymptotically, the Hodges–Lehmann (H-L) estimator is equivalent to 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) (23), Bickel and Freedman (24, 25), and Helmers, Janssen, and Veraverbeke (1990) (26)).

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

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

where $X_1 \leq \dots \leq X_n$ denote the order statistics of a sample of n independent and identically distributed random variables X_1, \dots, X_n , $\epsilon \bmod \frac{2}{b-1} = 0$, $\frac{1}{\epsilon} \geq 9$. If the subscript n is omitted, only the asymptotic behavior is considered. If b is omitted, $b = 3$ is assumed. The basic idea is to distribute the random variables into b blocks according to their order, and then compute the mean of the middle block, which is the median of all b blocks. Although the principle is similar to the median of means, without the random shift, the result is different from MoM $_{k=\frac{n}{b},b}$. The exact solution for $n \bmod \frac{b-1}{2\epsilon} \neq 0$ is imputing the remaining values with multiple hot deck imputation (proposed by Little and Rubin in 1986) (27), since it preserves the original distribution (proven by Reilly in 1991) (28). If $n \bmod \frac{b-1}{2\epsilon} = \varrho$, the algorithm should run $\binom{n}{\varrho}$ times. An approximation solution is randomly imputing the remaining values several times and then computing the mean of all estimations. The stratified mean is a type of stratum mean which is related to the stratified sampling. The most similar version was proposed by Takahasi and Wakimoto in 1968 (29), which is stratifying order statistics into several non-overlapping blocks and then computing the mean of one block. The median of means and stratified mean are consistent

Significance Statement

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

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

The author declares no competing interest.

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

mean estimators if their asymptotic breakdown points are zero. However, if $\epsilon = \frac{1}{9}$, the biases of the $SM_{\frac{1}{9}}$ are nearly identical to those of the $WM_{\frac{1}{9}}$ in asymmetric distributions (Figure ??, if no other subscripts, ϵ is omitted for simplicity), i.e., their robustness to departures from the symmetry assumption is similar in practice. More importantly, the bounds confirm that the worst-case performances of WM_{ϵ} are better than those of TM_{ϵ} in terms of bias, but due to the complexity, any extensions are extremely difficult. The aim of this paper is to define a series of semiparametric models using inequalities, demonstrate their elegant interrelations and connections to parametric models, and deduce a set of sophisticated robust mean estimators.

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

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

where w_i s are the weights applied to the symmetric quantile average 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 ($\lim_{\epsilon \rightarrow 0} TM_{\epsilon}$) and median ($TM_{\frac{1}{2}}$) are two special cases of symmetric trimmed mean. In 1974, Hogg investigated asymmetric trimmed mean and found its advantages for some special applications (30). To extend to the asymmetric case, the quantile average can be defined as

$$QA(\epsilon, \gamma, n) := \frac{1}{2}(\hat{Q}_n(\epsilon) + \hat{Q}_n(1 - \gamma\epsilon)).$$

where $\gamma > 0$ and $\epsilon \leq \frac{1}{1+\gamma}$, $\hat{Q}_n(p)$ is the empirical quantile function. For example, $QA(\epsilon = 0.2, \gamma = 2, n) = \frac{1}{2}(\hat{Q}_n(0.2) + \hat{Q}_n(0.6))$. Symmetric quantile average is a special case of quantile average when $\gamma = 1$.

Analogously, weighted average can be defined as

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

Converting this asymptotic definition to finite sample definition requires rounding the $n\epsilon_0$, for simplicity, only asymptotic definition is considered here. For example, the ϵ, γ -asymmetric trimmed mean ($TM_{\epsilon,\gamma}$) is a weighted average that trims the left

side ϵ and trims the right side $\gamma\epsilon$, 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, every quantile have their corresponding "pair" with the same γ . For the sake of brevity, in the following text, if γ is not indicated, symmetry will be assumed.

Classifying distributions through inequalities

Let \mathcal{P}_k denote the set of all distributions over \mathbb{R} whose moments, from the first to the k th, are all finite. Without loss

of generality, all classes discussed in the following are subclasses of the nonparametric class of distributions such that $\mathcal{P}_1^k := \{P \text{ is continuous} \wedge \text{all } P \in \mathcal{P}_k\}$. Besides fully and smoothly parameterized 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 foundational 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). They systematically classified nearly all 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 inequalities, 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.

There was a widespread misbelief that the median is always located between the mean and the mode for an arbitrary unimodal distribution until Runnenburg (1978) and van Zwet (1979) (33, 34) endeavored to determine sufficient conditions under which the inequality holds, thus implying the possibility of its violation (counterexamples see Dharmadhikari and Joag-Dev (1988), Basu and DasGupta (1997), and Abadir (2005)'s papers) (35–37). The class of distributions satisfying the mean-median-mode inequality constitutes a subclass of \mathcal{P}_U . Analogously, a right-skewed distribution is called γ -ordered, if and only if

$$\forall \epsilon_1 \leq \epsilon_2 \leq \frac{1}{1+\gamma}, QA_{\epsilon_1,\gamma} \geq QA_{\epsilon_2,\gamma}.$$

It is reasonable, although not necessary, to further assume $\gamma \geq 1$ since the gross errors of a right-skewed distribution, often, are mainly from the right side. The left-skewed case is just reversing the inequality and, if needed, assuming $\gamma \leq 1$; for simplicity, it will be completely omitted in the following discussion. If $\gamma = 1$, it is referred to as ordered. This nomenclature will be assumed in the following text. Let \mathcal{P}_O denote the set of all ordered distributions. Nearly all common unimodal distributions, including Weibull, gamma, Pareto, lognormal and generalized Gaussian, are in $\mathcal{P}_U \cap \mathcal{P}_O$ (proven in the following discussion and SI Text). The only minor exceptions occur when the Weibull and gamma distribution are near-symmetric (shown in the SI Text). Unlike the mean-median-mode inequality, whose sufficient conditions are very cumbersome, a necessary and sufficient condition of the γ -orderliness is the monotonic property of the bias function of $QA_{\epsilon,\gamma}$ with respect to ϵ (proven in the SI Text). The following necessary and sufficient condition hints at the relation between the mean-median-mode inequality and the orderliness.

Theorem .1. Let P_{Υ}^k denote an arbitrary distribution in the set \mathcal{P}_{Υ}^k . $P_{\Upsilon}^k \in \mathcal{P}_O$ if and only if the pdf satisfies the inequality $f(Q(\epsilon)) \geq f(Q(1 - \epsilon))$, where $0 \leq \epsilon \leq \frac{1}{2}$ (also assumed in the following discussions), $Q(\epsilon)$ is the quantile function.

Proof. From the definition of ordered distribution, deducing $\frac{Q(\epsilon - \delta) + Q(1 - \epsilon + \delta)}{2} \geq \frac{Q(\epsilon) + Q(1 - \epsilon)}{2} \Leftrightarrow Q(\epsilon - \delta) - Q(\epsilon) \geq Q(1 -$

142 $\epsilon) - Q(1 - \epsilon + \delta) \Leftrightarrow Q'(1 - \epsilon) \geq Q'(\epsilon)$, where δ is an infinitesimal
 143 quantity. Since the quantile function is the inverse function
 144 of the cumulative distribution function (cdf), $Q'(1 - \epsilon) \geq$
 145 $Q'(\epsilon) \Leftrightarrow F'(Q(\epsilon)) \geq F'(Q(1 - \epsilon))$, the proof is complete by
 146 noticing that the derivative of cdf is pdf. \square

147 The mean-median difference $|\mu - m|$ was proposed to mea-
 148 sure skewness by Pearson (1895) (38). Bowley (1926) pro-
 149 posed a robust skewness based on the SQA-median difference
 150 $|SQA_e - m|$ (39). Groeneveld and Meeden (1984) (40) gener-
 151 alized these measures of skewness based on van Zwet's convex
 152 transformation (41) and investigated their properties. Sup-
 153 pose P_T^k follows the mean-median-mode inequality. Then, the
 154 probability density $f(Q(\epsilon))$ on the left side of the median, on
 155 average, is greater than the corresponding $f(Q(1 - \epsilon))$, since
 156 $m < \frac{Q(0)+Q(1)}{2} \Leftrightarrow m - Q(0) < Q(1) - m$. If $Q(\epsilon) > M$, the
 157 inequality $f(Q(\epsilon)) \geq f(Q(1 - \epsilon))$ holds. The principle can be
 158 extended to unimodal-like distributions. Suppose there is a
 159 right-skewed continuous multimodal distribution following the
 160 mean-median-first mode inequality with many small modes on
 161 the right side, the first mode, M , has the greatest probability
 162 density and the median is within the first dominant mode,
 163 i.e., if $x > m$, $f(m) \geq f(x)$, then, if $Q(\epsilon) > M$, the inequality
 164 $f(Q(\epsilon)) \geq f(Q(1 - \epsilon))$ will also hold.

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

$$\forall \epsilon_1 \leq \epsilon_2 \leq \epsilon_3 \leq \frac{1}{1 + \gamma}, \frac{QA_{\epsilon_1, \gamma} - QA_{\epsilon_2, \gamma}}{\epsilon_2 - \epsilon_1} \geq \frac{QA_{\epsilon_2, \gamma} - QA_{\epsilon_3, \gamma}}{\epsilon_3 - \epsilon_2}.$$

165 An equivalent expression is $\frac{d^2 QA}{d\epsilon^2} \geq 0 \wedge \frac{dQA}{d\epsilon} \leq 0$. Analogously,
 166 the ν th γ -orderliness can be defined as $(-1)^\nu \frac{d^\nu QA}{d\epsilon^\nu} \geq 0 \wedge \dots \wedge$
 167 $-\frac{dQA}{d\epsilon} \geq 0$. The definition of ν th orderliness is the same, just
 168 setting $\gamma = 1$. Common unimodal distributions are also second
 169 and third ordered (shown in the SI Text). Let \mathcal{P}_{O_ν} and $\mathcal{P}_{\gamma O_\nu}$
 170 denote the sets of all distributions which are ν th ordered and
 171 ν th γ -ordered. The following theorems can be used to quickly
 172 identify parametric distributions in \mathcal{P}_{O_ν} and $\mathcal{P}_{\gamma O_\nu}$ without
 173 solving the exact derivative.

174 **Theorem .2.** Any symmetric distribution with a finite second
 175 moment is ν th ordered.

176 *Proof.* The assertion follows from the fact that for any sym-
 177 metric distribution with a finite second moment, all symmetric
 178 quantile averages coincide. Therefore, the SQA function is
 179 always a horizontal line; the ν th order derivative is zero. \square

180 As a consequence of Theorem .2 and the fact that general-
 181 ized Gaussian distribution is symmetric around the median, it
 182 is ν th ordered.

183 **Theorem .3.** Any continuous right skewed distribution whose
 184 Q satisfies $Q^{(\nu)}(p) \geq 0 \wedge \dots \wedge Q^{(i)}(p) \geq 0 \dots \wedge Q^{(2)}(p) \geq 0$,
 185 $i \bmod 2 = 0$, is ν th γ -ordered, provided that $\gamma \geq 1$.

186 *Proof.* Let $QA(\epsilon) = \frac{1}{2}(Q(\epsilon) + Q(1 - \gamma\epsilon))$, then $(-1)^j \frac{d^j QA}{d\epsilon^j} =$
 187 $\frac{1}{2}((-1)^j Q^j(\epsilon) + \gamma^j Q^j(1 - \gamma\epsilon))$, $\nu \geq j \geq 1$, when $j \bmod 2 = 0$,
 188 $(-1)^j \frac{d^j QA}{d\epsilon^j} \geq 0$, when $j \bmod 2 = 1$, the strict positivity
 189 is uncertain. If assuming $\gamma \geq 1$, $(-1)^j \frac{d^j QA}{d\epsilon^j} \geq 0$, since
 190 $Q^{(j+1)}(\epsilon) \geq 0$. \square

191 It is now trivial to prove that the Pareto distribution follows
 192 the ν th γ -orderliness, provided that $\gamma \geq 1$, since the quantile
 193 function of the Pareto distribution is $Q(p) = x_m(1 - p)^{-\frac{1}{\alpha}}$,
 194 $x_m > 0$, $\alpha > 0$, $Q^{(\nu)}(p) \geq 0$ according to the chain rule.

195 **Theorem .4.** A right-skewed continuous distribution with a
 196 monotonic decreasing pdf is γ -ordered, if $\gamma \geq 1$.

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 differenti-
 ating 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 Theo-
 rem .3. \square

203 Theorem .4 gives an interesting insight into the relation
 204 between modality and γ -orderliness. According to the conven-
 205 tional definition, a distribution with a monotonic pdf is still
 206 a unimodal distribution. However, within the interval sup-
 207 ported, its mode number is zero. In fact, the number of modes
 208 and their magnitudes are closely related to the possibility of
 209 the validity of orderliness, even though counterexamples can
 210 always be constructed. A proof of γ -orderliness, if $\gamma \geq 1$, can
 211 be easily done for the gamma distributions when $\alpha \leq 1$ since
 212 the pdf of the gamma distribution is $f(x) = \frac{\lambda^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\lambda}}}{\Gamma(\alpha)}$,
 213 $x \geq 0$, $\lambda > 0$, $\alpha > 0$, which is a product of two monotonic
 214 decreasing functions under constraints. For $\alpha > 1$, the proof
 215 is hard, numerical results show that the orderliness is valid
 216 until $\alpha > 140$ (SI Text), but it is instructive to consider that
 217 when $\alpha \rightarrow \infty$ the gamma distribution converges to a Gaussian
 218 distribution with mean $\mu = \alpha\lambda$ and variance $\sigma = \alpha\lambda^2$.

219 **Theorem .5.** If transforming a symmetric unimodal random
 220 variable X with a function $\phi(x)$ such that $\frac{d^2 \phi}{dx^2} \geq 0 \wedge \frac{d\phi}{dx} \geq$
 221 0 over the interval supported, then the convex transformed
 222 distribution is ordered. If the quantile function of X satisfies
 223 $Q^{(2)}(\epsilon) \leq 0$, the convex transformed distribution is second
 224 ordered.

225 *Proof.* Let $\phi SQA(\epsilon) = \frac{1}{2}(\phi(Q(\epsilon)) + \phi(Q(1 - \epsilon)))$, then,
 226 $\frac{d\phi SQA}{d\epsilon} = \frac{1}{2}(\phi'(Q(\epsilon))Q'(\epsilon) - \phi'(Q(1 - \epsilon))Q'(1 - \epsilon)) =$
 227 $\frac{1}{2}Q'(\epsilon)(\phi'(Q(\epsilon)) - \phi'(Q(1 - \epsilon))) \leq 0$, since for a symmet-
 228 ric distribution, $m - Q(\epsilon) = Q(1 - \epsilon) - m$, differentiat-
 229 ing both sides, $-Q'(\epsilon) = -Q'(1 - \epsilon)$, $Q'(\epsilon) \geq 0$, $\phi^{(2)} \geq$
 230 0. Notably, differentiating twice, $Q^{(2)}(\epsilon) = -Q^{(2)}(1 -$
 $\epsilon)$, $\frac{d^{(2)} \phi SQA}{d\epsilon^{(2)}} = \frac{1}{2}((\phi^{(2)}(Q(\epsilon)) + \phi^{(2)}(Q(1 - \epsilon)))(Q'(\epsilon))^2 +$
 $\frac{1}{2}((\phi'(Q(\epsilon)) - \phi'(Q(1 - \epsilon)))Q^{(2)}(\epsilon))$. The sign of $\frac{d^{(2)} \phi SQA}{d\epsilon^{(2)}}$
 depends on $Q^{(2)}(\epsilon)$. \square

234 The mean-median-mode inequality for distributions of the
 235 powers and roots of the variates of a given distribution was
 236 investigated by Henry Rietz in 1927 (42), but the most trivial
 237 solution is the exponential transformation since the deriva-
 238 tives are always positive. An application of Theorem .5
 239 is that the lognormal distribution is ordered as it is expo-
 240 nentially transformed from the Gaussian distribution whose
 241 $Q^{(2)}(\epsilon) = -2\sqrt{2\pi}\sigma e^{2\text{erfc}^{-1}(2\epsilon)^2} \text{erfc}^{-1}(2\epsilon) \leq 0$ (so, it is also
 242 second ordered).

243 Theorem .5 also reveals a relation between convex transfor-
 244 mation and orderliness, since ϕ is the non-decreasing convex

function in van Zwet's trailblazing work *Convex transformations of random variables* (41). Consider there is a near-symmetric distribution S such that SQA_ϵ as a function of ϵ is fluctuating from 0 to $\frac{1}{2}$, and $\mu = m$. Based on the definition, S is not ordered. Let s be the pdf of S . Transforming S with $\phi(x)$ will decrease $s(Q_S(\epsilon))$, and the decrease rate, due to the order, is much smaller than $s(Q_S(1-\epsilon))$. That means, as the second derivative of $\phi(x)$ increases, eventually, after a point, $s(Q_S(\epsilon))$ will always be greater than $s(Q_S(1-\epsilon))$ even previously not, i.e., the SQA_ϵ function will be monotonic decreasing and S will be 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.

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

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

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

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

When $\frac{1}{2} \geq \epsilon > \frac{1}{6}$, $B'_{\text{SQAB}}(\epsilon) = \frac{1}{(4-3\epsilon)^2 \sqrt{\frac{1-\epsilon}{9\epsilon+3}}} - \frac{1}{(3\epsilon+1)^2 \sqrt{\frac{1-\epsilon}{9\epsilon+3}}}$. To check whether $B'_{\text{SQAB}}(\epsilon) < 0$, 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, $(4-3\epsilon)^{\frac{3}{2}} \sqrt{\epsilon} > (3\epsilon+1)^{\frac{3}{2}} \sqrt{1-\epsilon} \sqrt{\frac{1}{3}} \iff (4-3\epsilon)^{\frac{3}{2}} \sqrt{\epsilon} > (3\epsilon+1)^{\frac{3}{2}} \sqrt{1-\epsilon} \sqrt{\frac{1}{3}} \iff 3(4-3\epsilon)^3 \epsilon > (3\epsilon+1)^3 (1-\epsilon) \iff -54\epsilon^4 + 324\epsilon^3 - 450\epsilon^2 + 184\epsilon - 1 > 0$. Since when $\epsilon < 1$, $-54\epsilon^4 + 54\epsilon^3 > 0$, just consider the condition that $270\epsilon^3 - 450\epsilon^2 + 184\epsilon - 1 > 0 \iff \epsilon(270\epsilon^2 - 450\epsilon + 174) + 10\epsilon - 1 > 0$. Since $270\epsilon^2 - 450\epsilon + 174 > 0$ is valid for $\epsilon < \frac{1}{30}(25 - 3\sqrt{5})$, so just need $10\epsilon - 1 > 0$, $10\epsilon > 1$, $\epsilon > \frac{1}{10}$. So, the inequality is valid for $\frac{1}{30}(25 - 3\sqrt{5}) \approx 0.610 > \epsilon > \frac{1}{10}$, within the range of $\frac{1}{2} \geq \epsilon > \frac{1}{6}$, therefore, $B'_{\text{SQAB}} < 0$ for $\frac{1}{2} \geq \epsilon > \frac{1}{6}$. The first and second formula, when $\epsilon = \frac{1}{6}$, are all

equal to $\frac{1}{2} \left(\sqrt{\frac{4}{9}} + \frac{1}{\sqrt{7}} \right)$. It follows that $B_{\text{SQAB}}(\epsilon)$ is continuous over $(0, \frac{1}{2})$. Hence, $B'_{\text{SQAB}}(\epsilon) < 0$ is valid for $0 < \epsilon < \frac{1}{2}$, which leads to the assertion of this theorem. \square

The proof is given in the SI Text. This monotonicity indicates that the extent of any violations of the orderliness is bounded for a unimodal distribution, e.g., for a right-skewed unimodal distribution, if $\exists \epsilon_1 \leq \epsilon_2 \leq \epsilon_3 \leq \frac{1}{2}$, $\text{SQA}_{\epsilon_2} \geq \text{SQA}_{\epsilon_3} \geq \text{SQA}_{\epsilon_1}$, SQA_{ϵ_2} will not be too far away from SQA_{ϵ_1} , since $\sup_{P \in \mathcal{P}_U \cap \mathcal{P}_T^2} (\text{SQA}_{\epsilon_1}) > \sup_{P \in \mathcal{P}_U \cap \mathcal{P}_T^2} (\text{SQA}_{\epsilon_2}) > \sup_{P \in \mathcal{P}_U \cap \mathcal{P}_T^2} (\text{SQA}_{\epsilon_3})$.

Inequalities related to weighted averages

The bias bound of the ϵ -symmetric trimmed mean is also monotonic for $\mathcal{P}_U \cap \mathcal{P}_2$ (proven in the SI Text using the formula provided in Bernard et al.'s paper) (2). So far, it appears clear that the average bias of an estimator is closely related to its degree of robustness. For a right-skewed unimodal distribution, often, the mean, ϵ -symmetric trimmed mean, and median occur in reverse alphabetical order, then analogous to the orderliness, $\forall \epsilon_1 \leq \epsilon_2 \leq \frac{1}{2}$, $\text{TM}_{\epsilon_1} \geq \text{TM}_{\epsilon_2}$. It is called trimming inequality. γ -trimming inequality is the same, just added the term γ . A necessary and sufficient condition of the γ -trimming inequality is also the monotonic behavior of the bias of the ϵ, γ -trimmed mean as a function of the breakdown point ϵ (proven in the SI Text). Orderliness is a sufficient condition for the trimming inequality (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 $\frac{1}{1-\epsilon-\gamma\epsilon+2\delta} \int_{\epsilon-\delta}^{1-\gamma\epsilon+\delta} Q(u) du \geq \frac{1}{1-\epsilon-\gamma\epsilon} \int_{\epsilon}^{1-\gamma\epsilon} Q(u) du \iff \int_{\epsilon-\delta}^{1-\gamma\epsilon+\delta} Q(u) du - \frac{1-\epsilon-\gamma\epsilon+2\delta}{1-\epsilon-\gamma\epsilon} \int_{\epsilon}^{1-\gamma\epsilon} Q(u) du \geq 0 \iff \int_{\epsilon-\delta}^{\epsilon} Q(u) du - \frac{2\delta}{1-\epsilon-\gamma\epsilon} \int_{\epsilon}^{1-\gamma\epsilon} Q(u) du \geq 0 \iff \frac{1}{2\delta} \left(\int_{\epsilon-\delta}^{1-\gamma\epsilon+\delta} Q(u) du + \int_{\epsilon}^{\epsilon} Q(u) du \right) \geq \frac{1}{1-\epsilon-\gamma\epsilon} \int_{\epsilon}^{1-\gamma\epsilon} Q(u) du$ and noting that $\lim_{\delta \rightarrow 0} \frac{1}{2\delta} \left(\int_{\epsilon-\delta}^{1-\gamma\epsilon+\delta} Q(u) du + \int_{\epsilon}^{\epsilon} Q(u) du \right) = \frac{Q(\epsilon) + Q(1-\gamma\epsilon)}{2}$, the proof is complete. \square

Additionally, the orderliness is a sufficient condition of the Winsorization inequality (proven in the SI Text). 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 $\gamma \geq 1$.

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

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 $\gamma \geq 1$. For simplicity, just consider the symmetric case, if one wants to construct a robust SWA following the mean-symmetric weighted average-median inequality. According to the definition of the orderliness, for a right-skewed or symmetric ordered distribution, $SWA \geq m$. So, the condition that SWA is between m and μ is $\mu \geq SWA$. Let $\mathcal{B}_i = \int_{i\epsilon}^{(i+1)\epsilon} SQA(u) du$, $ka = k\epsilon + b$, from the orderliness, $-\frac{dSQA}{d\epsilon} \geq 0 \Rightarrow \forall a \leq 2a \leq \frac{1}{2}, -\frac{(SQA(2a)-SQA(a))}{a} \geq 0 \Rightarrow \mathcal{B}_i - \mathcal{B}_{i+1} \geq 0$. Let $\mathcal{B}_i = \mathcal{B}_0$, then, from the orderliness, another version of ϵ -symmetric Winsorized mean, block Winsorized mean, is defined here for comparison in the SI Dataset S1 as

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

which is double weighting the leftest and rightest blocks having the size ϵn . From the second orderliness $\frac{d^2 SQA}{d^2 \epsilon} \geq 0 \Rightarrow \forall a \leq 2a \leq 3a \leq \frac{1}{2}, \frac{1}{a} \left(\frac{(SQA(3a)-SQA(2a))}{a} - \frac{(SQA(2a)-SQA(a))}{a} \right) \geq 0 \Rightarrow \mathcal{B}_i - 2\mathcal{B}_{i+1} + \mathcal{B}_{i+2} \geq 0$. So, from the second orderliness, SM_ϵ can be seen as replacing the two blocks, $\mathcal{B}_i + \mathcal{B}_{i+2}$ with one block $2\mathcal{B}_{i+1}$. From the ν th orderliness, the recurrence relation of the derivatives naturally produces the alternating binomial coefficients,

$$\begin{aligned} (-1)^\nu \frac{d^\nu SQA}{d\epsilon^\nu} &\geq 0 \Rightarrow \forall a \leq \dots \leq (\nu+1)a \leq \frac{1}{2}, \\ \frac{(-1)^\nu}{a} \left(\frac{SQA(\nu a + a)}{a} - \frac{SQA(2a)}{a} - \frac{SQA(\nu a)}{a} + \frac{SQA(a)}{a} \right) \\ &\geq 0 \Leftrightarrow \frac{(-1)^\nu}{a^\nu} \left(\sum_{j=0}^{\nu} (-1)^j \binom{\nu}{j} SQA((\nu-j+1)a) \right) \geq 0 \\ &\Rightarrow \sum_{j=0}^{\nu} (-1)^j \binom{\nu}{j} \mathcal{B}_{i+j} \geq 0. \end{aligned}$$

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

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

where $\mathfrak{B}_{i,j} = \sum_{l=n\epsilon(j+(i-1)(\nu+1)+1)}^{n\epsilon(j+(i-1)(\nu+1)+1)} (X_l + X_{n-l+1})$, $\frac{1}{2}\epsilon^{-1}(\nu+1)-1 \in \mathbb{N}$. If ν is not indicated, it is default as $\frac{1}{2}\epsilon^{-1}-1$. The solution for $n \bmod \epsilon^{-1} \neq 0$ is the same as that in the stratified mean. $BM_{\nu=2,\epsilon}$ is identical to $SM_{\epsilon,b=3}$, if the two ϵ s are the same. The reason that $SM_{\frac{1}{9}}$ has similar biases as the corresponding $WM_{\frac{1}{9}}$ is the Winsorized mean is using two single quantiles to replace the trimmed parts, not two blocks. If using single quantiles, based on the second orderliness, the stratified quantile mean can be defined as

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

where $\frac{1}{\epsilon} \bmod 4 = 0$, $\hat{Q}_n(p)$ is the empirical quantile function. $SQM_{\frac{1}{4}}$ is the Tukey's midhinge (43). The biases of $SQM_{\frac{1}{8}}$ are also very close to those of $BM_{\frac{1}{8}}$ (Figure ??), which is based on the third orderliness.

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. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
11. F Wilcoxon, Individual comparisons by ranking methods. *Biom. Bull.* **1**, 80–83 (1945).
12. AS Nemirovskij, DB Yudin, *Problem complexity and method efficiency in optimization*. (Wiley-Interscience), (1983).
13. MR Jerrum, LG Valiant, VV Vazirani, Random generation of combinatorial structures from a uniform distribution. *Theor. computer science* **43**, 169–188 (1986).
14. 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).
15. 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).
16. JY Audibert, O Catoni, Robust linear least squares regression. *The Annals Stat.* **39**, 2766–2794 (2011).
17. D Hsu, S Sabato, Heavy-tailed regression with a generalized median-of-means in *International Conference on Machine Learning*. (PMLR), pp. 37–45 (2014).
18. S Minsker, Geometric median and robust estimation in banach spaces. *Bernoulli* **21**, 2308–2335 (2015).
19. C Brownlees, E Joly, G Lugosi, Empirical risk minimization for heavy-tailed losses. *The Annals Stat.* **43**, 2507–2536 (2015).
20. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.* **44**, 2695–2725 (2016).
21. E Joly, G Lugosi, Robust estimation of u-statistics. *Stoch. Process. their Appl.* **126**, 3760–3773 (2016).
22. P Laforge, S Cléménçon, P Bertail, On medians of (randomized) pairwise means in *International Conference on Machine Learning*. (PMLR), pp. 1272–1281 (2019).
23. B Efron, Bootstrap methods: Another look at the jackknife. *The Annals Stat.* **7**, 1–26 (1979).
24. PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. *The annals statistics* **9**, 1196–1217 (1981).
25. PJ Bickel, DA Freedman, Asymptotic normality and the bootstrap in stratified sampling. *The annals statistics* **12**, 470–482 (1984).
26. R Helmers, P Janssen, N Veraverbeke, *Bootstrapping U-quantiles*. (CWI. Department of Operations Research, Statistics, and System Theory [BS]), (1990).
27. RJ Little, DB Rubin, *Statistical analysis with missing data*. (John Wiley & Sons) Vol. 793, (2019).
28. M Reilly, *Semi-parametric methods of dealing with missing or surrogate covariate data*. (University of Washington), (1991).
29. K Takahasi, K Wakimoto, On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals institute statistical mathematics* **20**, 1–31 (1968).
30. RV Hogg, Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *J. Am. Stat. Assoc.* **69**, 909–923 (1974).
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. JT Runnenburg, Mean, median, mode. *Stat. Neerlandica* **32**, 73–79 (1978).
34. Wv Zwet, Mean, median, mode ii. *Stat. Neerlandica* **33**, 1–5 (1979).
35. S Basu, A DasGupta, The mean, median, and mode of unimodal distributions: a characterization. *Theory Probab. & Its Appl.* **41**, 210–223 (1997).
36. S Dharmadhikari, K Joag-Dev, *Unimodality, convexity, and applications*. (Elsevier), (1988).

- 430 37. KM Abadir, The mean-median-mode inequality: Counterexamples. *Econom. Theory* **21**,
431 477–482 (2005).
- 432 38. K Pearson, X. contributions to the mathematical theory of evolution.—ii. skew variation in
433 homogeneous material. *Philos. Transactions Royal Soc. London.(A.)* **186**, 343–414 (1895).
- 434 39. AL Bowley, *Elements of statistics*. (King) No. 8, (1926).
- 435 40. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D*
436 *(The Stat.* **33**, 391–399 (1984).
- 437 41. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen*. (1964).
- 438 42. H Rietz, On certain properties of frequency distributions of the powers and roots of the variates
439 of a given distribution. *Proc. Natl. Acad. Sci.* **13**, 817–820 (1927).
- 440 43. JW Tukey, , et al., *Exploratory data analysis*. (Reading, MA) Vol. 2, (1977).

DRAFT