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 February 27, 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 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 is 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 νth U-orderliness, binomial Hodges–Lehmann mean is proposed as the bias-optimum semiparametric robust mean estimator.

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_ϵ) for unimodal distributions, building on the work 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, which are 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 a 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 deduced the median of pairwise means as a robust location estimator for a symmetric population based on the Wilcoxon signed-rank statistic (11). The concept of median of means (MoM) 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. Devroye, Lerasle, Lugosi, and Oliveira (2016) showed that MoM nears the optimum of mean estimation with regards to concentration bounds when the distribution has a heavy tail (15). Here, the ϵ-stratified mean is defined as

$$\mathrm{SM}_{\epsilon,n} := \frac{3}{n}\left(\sum_{j=1}^{\frac{1}{3\epsilon}} \sum_{i_j=(3j-2)n\epsilon+1}^{(3j-1)n\epsilon} X_{i_j}\right),$$

where $X_1 \leq \ldots \leq X_n$ denote the order statistics of a sample of n independent and identically distributed random variables $X_1, \ldots, X_n$, $\frac{1}{\epsilon} \bmod 3 = 0$, $\frac{1}{\epsilon} \geq 9$. If the subscript n is omitted, only the asymptotic behavior is considered. The basic idea is to divide the random variables into three blocks according to their order, and then compute the mean of the middle block, which is the median of all three blocks. Thus, it is essentially a deterministic version of MoM. The exact solution for $n \bmod \frac{1}{\epsilon} \neq 0$ is imputing the remaining values with multiple hot deck imputation (proposed by Little and Rubin in 1986) (16), since it preserves the original distribution (proven by Reilly in 1991) (17). If $n \bmod \frac{1}{\epsilon} = \varrho$, 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. Having good performance even for distributions with infinite second moments, the advantages of MoM have received increasing attention over the past decade (15, 18–24). Most of these results should be similar for SM_ϵ. Using a stochastic approach, the variance of MoM is very high, SM_ϵ is obviously almost always a better choice (the only apparent drawback is its computational complexity is $O(n \log n)$, not $O(n)$). In fact, 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 (25), 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 mean estimators if their asymptotic breakdown points are zero. However, if $\epsilon = \frac{1}{9}$, the biases of the $\mathrm{SM}_{\frac{1}{9}}$ are nearly identical to those of the $\mathrm{WM}_{\frac{1}{9}}$ 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

Significance Statement

In 1964, van Zwet introduced the convex transformation order for comparing the skewness of two distributions. This paradigm shift plays a fundamental role in defining robust measures of distributions, from spread to kurtosis. Here, rather than the stochastic orderings 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 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

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

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

where w_is are the weights applied to the symmetric quantile 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 ($\lim_{\epsilon \to 0} \mathrm{TM}_\epsilon$) and median ($\mathrm{TM}_{\frac{1}{2}}$) are just two special cases of symmetric trimmed mean. In 1974, Hogg investigated asymmetric trimmed mean and found its advantages for some special applications (26). To extend to the asymmetric case, the quantile average can be defined as

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

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

Analogously, weighted average can be defined as

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

Converting this asymptotic definition to finite sample definition requires rounding the $n\epsilon_0$, for simplicity, only asymptotic definition is considered here. For example, the ϵ, γ-asymmetric trimmed mean ($\mathrm{TM}_{\epsilon,\gamma}$) 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 all quantile averages have the same γ in a weighted average. 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 kth, are all finite. Without loss of generality, all classes discussed in the following are subclasses of the nonparametric class of distributions such that $\mathcal{P}_\Upsilon^k := \{P \text{ is continuous } \wedge \text{ all } P \in \mathcal{P}_k\}$. 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 (27). A notable example discussed in his seminal 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 (28). 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) (29, 30) 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) (31–33). 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

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

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 $\mathrm{QA}_{\epsilon,\gamma}$ 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 $\mathcal{P}_\Upsilon^k$. $P_\Upsilon^k \in \mathcal{P}_O$ 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 $\frac{Q(\epsilon-\delta)+Q(1-\epsilon+\delta)}{2} \geq \frac{Q(\epsilon)+Q(1-\epsilon)}{2} \Leftrightarrow Q(\epsilon-\delta) - Q(\epsilon) \geq Q(1-\epsilon) - Q(1-\epsilon+\delta) \Leftrightarrow Q'(1-\epsilon) \geq Q'(\epsilon)$, where δ is an infinitesimal quantity. Since the quantile function is the inverse function of the cumulative distribution function (cdf), $Q'(1-\epsilon) \geq Q'(\epsilon) \Leftrightarrow F'(Q(\epsilon)) \geq F'(Q(1-\epsilon))$, the proof is complete by noticing that the derivative of cdf is pdf. □

The mean-median difference $|\mu - m|$ was proposed to measure skewness by Pearson (1895) (34). Bowley (1926) proposed a robust skewness based on the SQA-median difference $|\mathrm{SQA}_{\epsilon} - m|$ (35). Groeneveld and Meeden (1984) (36) generalized these measures of skewness based on van Zwet's convex transformation (37) and investigated their properties. Suppose P_{Υ}^{k} follows the mean-median-mode inequality. Then, the probability density $f(Q(\epsilon))$ on the left side of the median, on average, is greater than the corresponding $f(Q(1-\epsilon))$, since $m < \frac{Q(0)+Q(1)}{2} \Leftrightarrow m - Q(0) < Q(1) - m$. If $Q(\epsilon) > M$, the inequality $f(Q(\epsilon)) \geq f(Q(1-\epsilon))$ holds. The principle can be extended to unimodal-like distributions. Suppose there is a right-skewed continuous multimodal distribution following the mean-median-first mode inequality with many small modes on the right side, the first mode, M, has the greatest probability density and the median is within the first dominant mode, i.e., if $x > m$, $f(m) \geq f(x)$, then, if $Q(\epsilon) > M$, the inequality $f(Q(\epsilon)) \geq f(Q(1-\epsilon))$ will also holds.

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

$$\forall \epsilon_1 \leq \epsilon_2 \leq \epsilon_3 \leq \frac{1}{2}, \frac{\mathrm{QA}_{\epsilon_1,\gamma} - \mathrm{QA}_{\epsilon_2,\gamma}}{\epsilon_2 - \epsilon_1} \geq \frac{\mathrm{QA}_{\epsilon_2,\gamma} - \mathrm{QA}_{\epsilon_3,\gamma}}{\epsilon_3 - \epsilon_2}.$$

An equivalent expression is $\frac{d^2\mathrm{QA}}{d\epsilon^2} \geq 0 \wedge \frac{d\mathrm{QA}}{d\epsilon} \leq 0$. Analogously, the γ-νth 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 (shown in the SI Text). Let $\mathcal{P}_{O_{\nu}}$ and $\mathcal{P}_{\gamma O_{\nu}}$ denote the sets of all distributions which are νth ordered and γ-νth ordered. The following four theorems can be used to quickly identify parametric distributions in $\mathcal{P}_{O_{\nu}}$ and $\mathcal{P}_{\gamma O_{\nu}}$ without solving the exact derivative.

Theorem .2. *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. $\square$

As a consequence of Theorem .2 and the fact that generalized Gaussian distribution is symmetric around the median, it is νth ordered.

Theorem .3. *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 γ-νth ordered, provided that $\gamma \geq 1$.*

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

It is now trivial to prove that the Pareto distribution follows the γ-νth orderliness, provided that $\gamma \geq 1$, 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 .4. *A right-skewed continuous distribution with a 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 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 .3. $\square$

Theorem .4 gives an interesting insight into the relation between modality and γ-orderliness. According to the conventional definition, a distribution with a monotonic pdf is still a unimodal distribution. However, within the interval supported, its mode number is zero. In fact, the number of modes and their magnitudes are closely related to the possibility of the validity of orderliness, even though counterexamples can always be constructed. A proof of γ-orderliness, if $\gamma \geq 1$, can be easily done for the gamma distributions when $\alpha \leq 1$ since 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 hard, 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 .5. *If transforming a symmetric unimodal random variable X with a function $\chi(x)$ such that $\frac{d^2\chi}{dx^2} \geq 0 \wedge \frac{d\chi}{dx} \geq 0$ over the interval supported, then the convex transformed distribution is ordered.*

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

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 (38), but the most trivial solution is the exponential transformation since the derivatives are always positive. An application of Theorem .5 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 .5 also reveals an interesting relation between convex transformation and orderliness. Consider two distribution F_X and F_Y, $Y = \chi(X)$, then $F_Y(y) = F_X\left(\chi^{-1}(y)\right)$, so $F_Y^{-1} \circ F_X = \chi\left(F_X^{-1}(F_X(x))\right) = \chi(x)$. If removing the monotonicity constraint, this is equivalent to the convex transformation (37). Consider there is a near-symmetric distribution S such that SQA_{ϵ} as a function of ϵ is monotonic increasing from 0 to u and monotonic decreasing from u to $\frac{1}{2}$, and $\mu = m$. Based on the definition, S is not ordered. Then, the monotonic increase from 0 to u implies that $Q'(\epsilon) \geq Q'(1-\epsilon) \Leftrightarrow f(Q(1-\epsilon)) \geq f(Q(\epsilon))$ always holds for $0 \leq \epsilon \leq u$. Similarly, $Q'(\epsilon) \leq Q'(1-\epsilon) \Leftrightarrow f(Q(\epsilon)) \geq f(Q(1-\epsilon))$ always holds for

$u \leq \epsilon \leq \frac{1}{2}$. Transforming S with $\chi(x)$ will decrease $f(Q(\epsilon))$, and the decrease rate, due to the order, is much smaller than $f(Q(1-\epsilon))$. That means, as the second derivative of $\chi(x)$ increases, eventually, after a point, $f(Q(\epsilon))$ will always be greater than $f(Q(1-\epsilon))$, i.e., the SQA_ϵ function will be monotonic decreasing and S will eventually be ordered. Accordingly, in a family of distributions that differ by a skewness-increasing transformation satisfies $\frac{d^2\chi}{dx^2} \geq 0 \wedge \frac{d\chi}{dx} \geq 0$, the violation of orderliness often only occurs when the distribution is near-symmetric.

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

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

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

Proof. When $\frac{1}{6} \geq \epsilon \geq 0$, $B'_{\mathrm{SQAB}}(\epsilon) = \frac{1}{(4-3\epsilon)^2\sqrt{\frac{\epsilon}{12-9\epsilon}}} - \frac{1}{3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}}$. 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, $\frac{1}{\epsilon} \in \left(\frac{9}{4}, \infty\right)$, $\frac{L(\epsilon)}{\epsilon^2} > \frac{R(\epsilon)}{\epsilon^2} \iff \left(\frac{4}{\epsilon}-3\right)^2\sqrt{\frac{1}{\frac{12}{\epsilon}-9}} > 3\sqrt{\frac{4}{\epsilon}-9} \iff \left(\frac{4}{\epsilon}-3\right)^2 > 3\sqrt{\frac{4}{\epsilon}-9}\sqrt{\frac{12}{\epsilon}-9}$. 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 $\frac{1}{2} \geq \epsilon > \frac{1}{6}$, $B'_{\mathrm{SQAB}}(\epsilon) = \frac{1}{(4-3\epsilon)^2\sqrt{\frac{\epsilon}{12-9\epsilon}}} - \frac{1}{(3\epsilon+1)^2\sqrt{\frac{1-\epsilon}{9\epsilon+3}}}$. 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, $(4-3\epsilon)^{\frac{3}{2}}\sqrt{\epsilon} > (3\epsilon+1)^{\frac{3}{2}}\sqrt{1-\epsilon}\sqrt{\frac{1}{3}} \iff (4-3\epsilon)^{\frac{3}{2}}\sqrt{\epsilon} > (3\epsilon+1)^{\frac{3}{2}}\sqrt{1-\epsilon}\sqrt{\frac{1}{3}} \iff 3(4-3\epsilon)^3\epsilon > (3\epsilon+1)^3(1-\epsilon) \iff -54\epsilon^4 + 324\epsilon^3 - 450\epsilon^2 + 184\epsilon - 1 > 0$. Since when $\epsilon < 1$, $-54\epsilon^4+54\epsilon^3 > 0$, just consider the condition that $270\epsilon^3 - 450\epsilon^2 + 184\epsilon - 1 > 0 \iff \epsilon\left(270\epsilon^2 - 450\epsilon + 174\right) + 10\epsilon - 1 > 0$. 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 $\frac{1}{2} \geq \epsilon > \frac{1}{6}$, therefore, $B'_{\mathrm{SQAB}} < 0$ for $\frac{1}{2} \geq \epsilon > \frac{1}{6}$. The first and second formula, when $\epsilon = \frac{1}{6}$, are all equal to $\frac{1}{2}\left(\sqrt{\frac{5}{3}} + \frac{1}{\sqrt{7}}\right)$. 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. □

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 $\exists \epsilon_1 \leq \epsilon_2 \leq \epsilon_3 \leq \frac{1}{2}$, $\mathrm{SQA}_{\epsilon_2} \geq \mathrm{SQA}_{\epsilon_3} \geq \mathrm{SQA}_{\epsilon_1}$, $\mathrm{SQA}_{\epsilon_2}$ will not be too far away from $\mathrm{SQA}_{\epsilon_1}$, since $\sup_{P \in \mathcal{P}_U \cap \mathcal{P}^2_\Upsilon}(\mathrm{SQA}_{\epsilon_1}) > \sup_{P \in \mathcal{P}_U \cap \mathcal{P}^2_\Upsilon}(\mathrm{SQA}_{\epsilon_2}) > \sup_{P \in \mathcal{P}_U \cap \mathcal{P}^2_\Upsilon}(\mathrm{SQA}_{\epsilon_3})$.

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. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.* **44**, 2695–2725 (2016).
16. RJ Little, DB Rubin, *Statistical analysis with missing data.* (John Wiley & Sons) Vol. 793, (2019).
17. M Reilly, *Semi-parametric methods of dealing with missing or surrogate covariate data.* (University of Washington), (1991).
18. 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).
19. JY Audibert, O Catoni, Robust linear least squares regression. *The Annals Stat.* **39**, 2766–2794 (2011).
20. D Hsu, S Sabato, Heavy-tailed regression with a generalized median-of-means in *International Conference on Machine Learning.* (PMLR), pp. 37–45 (2014).
21. S Minsker, Geometric median and robust estimation in banach spaces. *Bernoulli* **21**, 2308–2335 (2015).
22. C Brownlees, E Joly, G Lugosi, Empirical risk minimization for heavy-tailed losses. *The Annals Stat.* **43**, 2507–2536 (2015).
23. E Joly, G Lugosi, Robust estimation of u-statistics. *Stoch. Process. their Appl.* **126**, 3760–3773 (2016).
24. P Laforgue, S Clémençon, P Bertail, On medians of (randomized) pairwise means in *International Conference on Machine Learning.* (PMLR), pp. 1272–1281 (2019).
25. 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).
26. RV Hogg, Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *J. Am. Stat. Assoc.* **69**, 909–923 (1974).
27. 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).
28. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparametric models.* (Springer) Vol. 4, (1993).
29. JT Runnenburg, Mean, median, mode. *Stat. Neerlandica* **32**, 73–79 (1978).
30. Wv Zwet, Mean, median, mode ii. *Stat. Neerlandica* **33**, 1–5 (1979).
31. S Basu, A DasGupta, The mean, median, and mode of unimodal distributions: a characterization. *Theory Probab. & Its Appl.* **41**, 210–223 (1997).
32. S Dharmadhikari, K Joag-Dev, *Unimodality, convexity, and applications.* (Elsevier), (1988).
33. KM Abadir, The mean-median-mode inequality: Counterexamples. *Econom. Theory* **21**, 477–482 (2005).
34. 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).
35. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).

36. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).
37. W van Zwet, Convex transformations: A new approach to skewness and kurtosis in *Selected Works of Willem van Zwet.* (Springer), pp. 3–11 (2012).
38. 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).