

Semiparametric mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on May 19, 2023

As one of the most fundamental problems in statistics, robust location estimation has many prominent solutions, such as the symmetric trimmed mean, symmetric Winsorized mean, Hodges–Lehmann estimator, Huber M -estimator, and median of means. Recent studies suggest that their maximum biases concerning the mean can be quite different in asymmetric distributions, but the underlying mechanisms and average performance remain largely unclear. In this article, similar to the mean-median-mode inequality, it is proven that 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 typically have smaller biases compared to the trimmed mean. Building on the U -orderliness, the superiority of the median Hodges–Lehmann mean is discussed.

semiparametric | mean-median-mode inequality | asymptotic | unimodal
| Hodges–Lehmann estimator

In 1823, Gauss (1) proved that for any unimodal distribution with a finite second moment, $|m - \mu| \leq \sqrt{\frac{3}{4}}\omega$, where μ is the population mean, m is the population median, and ω is the root mean square deviation from the mode, M . This pioneering work revealed that despite potential bias with respect to the mean in robust estimates, the deviation remains bounded in unit of a scale parameter under certain assumptions. Bernard, Kazzi, and Vanduffel (2020) (2) further derived asymptotic bias bounds of any quantile for unimodal distributions with finite second moments by reducing this optimization problem to a parametric one, which can be solved analytically. They showed that the population median, m , has the smallest maximum distance to the population mean, μ , among all symmetric quantile averages (SQA $_e$). Daniell, in 1920, (3) analyzed a class of estimators, linear combinations of order statistics, and identified that ϵ -symmetric trimmed mean (STM $_e$) belongs to this class. Another popular choice, the ϵ -symmetric Winsorized mean (SWM $_e$), named after Winsor and introduced by Tukey (4) and Dixon (5) in 1960, is also an L -estimator. Bieniek (2016) derived exact bias upper bounds of the Winsorized mean based on Danielak and Rychlik's work (2003) on the trimmed mean for any distribution with a finite second moment and confirmed that the former is smaller than the latter (6, 7). In 1963, Hodges and Lehmann (8) proposed a class of nonparametric location estimators based on rank tests and, from the Wilcoxon signed-rank statistic (9), deduced the median of pairwise means as a robust location estimator for a symmetric population. Both L -statistics and R -statistics achieve robustness essentially by removing a certain proportion of extreme values. In 1964, Huber (10) generalized maximum likelihood estimation to the minimization of the sum of a specific loss function, which measures the residuals between the data points and the model's parameters. Some L -estimators are also M -estimators, e.g., the sample mean is an M -estimator with a squared error loss function, while the sample median is

an M -estimator with an absolute error loss function (10). The Huber M -estimator is obtained by applying the Huber loss function that combines elements of both squared error and absolute error to achieve robustness against gross errors and high efficiency for contaminated Gaussian distributions (10). Sun, Zhou, and Fan (2020) examined the concentration bounds of Huber M -estimator (11). Mathieu (2022) (12) further derived the concentration bounds of M -estimators and demonstrated that, by selecting the tuning parameter which depends on the variance, Huber M -estimator can also be a sub-Gaussian estimator. The concept of median of means (MoM $_{k,b} = \frac{n}{k}$, k is the number of size in each block, b is the number of blocks) was implicitly introduced several times in Nemirovsky and Yudin (1983) (13), Jerrum, Valiant, and Vazirani (1986), (14) and Alon, Matias and Szegedy (1996) (15)'s works. Given its good performance even for distributions with infinite second moments, MoM has received increasing attention over the past decade (16–23). Devroye, Lerasle, Lugosi, and Oliveira (2016) showed that MoM nears the optimum of sub-Gaussian mean estimation with regards to concentration bounds when the distribution has a heavy tail (21). For a comparison of concentration bounds of trimmed mean, Huber M -estimator, median of means and other relevant estimators, readers are directed to Gobet, Lerasle, and Métivier's paper (2022) (24). Laforgue, Clemencon, and Bertail (2019) proposed the median of randomized means (MoRM $_{k,b}$) (23), wherein, rather than partitioning, an arbitrary number, b , of blocks are built independently from the sample, and showed that MoRM has better non-asymptotic sub-Gaussian property compared to MoM. In fact, asymptotically, the Hodges–Lehmann (H-L) estimator is equivalent to MoM $_{k=2,b=\frac{n}{k}}$ and MoRM $_{k=2,b}$, and they can be seen as the pairwise mean distribution is approximated by the sampling without replacement and bootstrap, respectively. For the asymptotic validity, readers are referred to the foundational works of Efron (1979) (25), Bickel and Freedman (1981,

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.

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

1984) (26, 27), and Helmers, Janssen, and Veraverbeke (1990) (28).

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

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

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

Quantile average and weighted average

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

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

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

To extend the symmetric quantile average to the asymmetric case, there are two possible definitions for the ϵ, γ -quantile average ($\text{QA}(\epsilon, \gamma, n)$), i.e.,

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

and

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

where $\gamma \geq 0$ and $0 \leq \epsilon \leq \frac{1}{1+\gamma}$, $\hat{Q}_n(p)$ is the empirical quantile function. For trimming from both sides, [1] and [2] are equivalent. [1] is assumed in this article unless otherwise specified, since many common asymmetric distributions are right skewed, and [1] allows trimming only from the right side by setting $\gamma = 0$.

Analogously, the weighted average can be defined as

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

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

Classifying distributions by the signs of derivatives

Let \mathcal{P}_k denote the set of all distributions over \mathbb{R} whose moments, from the first to the k th, are all finite. Without loss of generality, all classes discussed in the following are subclasses of the nonparametric class of distributions $\mathcal{P}_\infty^k := \{\text{All continuous distribution } P \in \mathcal{P}_k\}$. Besides fully and smoothly parameterizing by a Euclidean parameter or just assuming regularity conditions, there are many ways to classify distributions. In 1956, Stein initiated the problem of estimating parameters in the presence of an infinite dimensional nuisance shape parameter (31). A notable example discussed in his groundbreaking work was the estimation of the center of symmetry for an unknown symmetric distribution. In 1993, Bickel, Klaassen, Ritov, and Wellner published an influential semiparametrics textbook (32) and systematically classified many common models into three classes: parametric, nonparametric, and semiparametric. However, there is another old and commonly encountered class of distributions that receives little attention in semiparametric literature: the unimodal distribution. It is a very unique semiparametric model because its definition is based on the signs of derivatives, i.e., assuming P is continuous, $(f'(x) > 0 \text{ for } x \leq M) \wedge (f'(x) < 0 \text{ for } x \geq M)$. Let \mathcal{P}_U denote the set of all unimodal distributions. 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) (33, 34) 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)) (35–37). The class of distributions that satisfy 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

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

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

Theorem .1. Let P_γ^k denote an arbitrary distribution in the set \mathcal{P}_γ^k . $P_\gamma^k \in \mathcal{P}_{\gamma O}$ if and only if the pdf satisfies the inequality $f(Q(\gamma\epsilon)) \geq f(Q(1-\epsilon))$ for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ or $f(Q(\gamma\epsilon)) \leq f(Q(1-\epsilon))$ for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$, $\gamma \geq 0$.

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

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

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

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

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

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

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

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

Analogously, the ν th γ -orderliness of a right-skewed distribution can be defined as $(-1)^\nu \frac{\partial^\nu \text{QA}_{\epsilon, \gamma}}{\partial \epsilon^\nu} \geq 0 \wedge \dots \wedge -\frac{\partial \text{QA}_{\epsilon, \gamma}}{\partial \epsilon} \geq 0$. If $\gamma = 1$, the ν th γ -orderliness is referred as ν th orderliness. Let \mathcal{P}_O denote the set of all distributions that are ordered and let \mathcal{P}_{O_ν} and $\mathcal{P}_{\gamma O_\nu}$ denote the sets of all distributions that are ν th ordered and ν th γ -ordered, respectively. Five parametric distributions in \mathcal{P}_U are detailed as examples here: Weibull, gamma, Pareto, lognormal and generalized Gaussian. When the shape parameters of the Weibull distributions fall within a certain range, it can be shown that it belong to $\mathcal{P}_U \cap \mathcal{P}_O \cap \mathcal{P}_{O_2} \cap \mathcal{P}_{O_3}$ (SI Text). The parameters that let it not be included in the set correspond to cases when it is near-symmetric, as shown in the SI Text. Nevertheless, computing the derivatives of the QA function is often cumbersome and, at times, challenging. The following theorems can be used to quickly identify parametric distributions in \mathcal{P}_O , \mathcal{P}_{O_ν} , and $\mathcal{P}_{\gamma O_\nu}$.

Theorem .2. For any random variable X whose probability distribution function belongs to a location-scale family, the distribution is ν th γ -ordered if and only if the family of probability distributions is ν th γ -ordered.

Proof. Let Q_0 denote the quantile function of the standard distribution without any shifts or scaling, then, after a location-scale transformation, the quantile function is $Q(p) = \lambda Q_0(p) + \mu$, where λ is the scale parameter, μ is the location parameter. According to the definition of the ν th γ -orderliness, the signs of derivatives of the QA function remain the same after this transformation. Since the location-scale transformation can also be performed inversely, the proof is complete. \square

Theorem .2 shows that in the analytical proof of the ν th γ -orderliness of a parametric distribution, both the location and scale parameters can be regarded as constants. It is also instrumental in proving other theorems, as illustrated below.

Theorem .3. Any symmetric distribution with a finite second moment is ν th ordered.

Proof. Without loss of generality, assuming continuity and $m = 0$, a symmetric distribution is a probability distribution such that for all x , $f(x) = f(-x)$. The cdf of it satisfies $F(x) = 1 - F(-x)$. Let $x = Q(p)$, then, $F(Q(p)) = p = 1 - F(-Q(p))$ and $F(Q(1-p)) = 1 - p \Leftrightarrow p = 1 - F(Q(1-p))$. Therefore, $F(-Q(p)) = F(Q(1-p))$. Since the cdf is monotonic, $-Q(p) = Q(1-p) \Leftrightarrow Q(p) + Q(1-p) = 0$. As a result, all symmetric quantile averages coincide; the ν th order derivative is zero. The case of $m \neq 0$ follows directly from Theorem .2. \square

As a consequence of Theorem .3 and the fact that generalized Gaussian distribution is symmetric around the median, it is ν th ordered.

Theorem .4. Any continuous right skewed distribution whose quantile function Q satisfies $Q^{(\nu)}(p) \geq 0 \wedge \dots \wedge Q^{(i)}(p) \geq 0 \wedge \dots \wedge Q^{(2)}(p) \geq 0$, $i \bmod 2 = 0$, is ν th γ -ordered, provided that $0 \leq \gamma \leq 1$.

230 *Proof.* Since $(-1)^i \frac{\partial^i \text{QA}_{\epsilon, \gamma}}{\partial \epsilon^i} = \frac{1}{2}((- \gamma)^i Q^i(\gamma \epsilon) + Q^i(1 - \epsilon))$, $0 \leq$
 231 $\epsilon \leq \frac{1}{1+\gamma}$, $1 \leq i \leq \nu$, when $i \bmod 2 = 0$, $(-1)^i \frac{\partial^i \text{QA}_{\epsilon, \gamma}}{\partial \epsilon^i} \geq 0$ for
 232 all $\gamma \geq 0$. When $i \bmod 2 = 1$, if further assuming $0 \leq \gamma \leq 1$,
 233 $(-1)^i \frac{\partial^i \text{QA}_{\epsilon, \gamma}}{\partial \epsilon^i} \geq 0$, since $Q^{(i+1)}(\epsilon) \geq 0$. \square

234 It is now trivial to prove that the Pareto distribution follows
 235 the ν th γ -orderliness, provided that $0 \leq \gamma \leq 1$, since the
 236 quantile function of the Pareto distribution is $Q(p) = x_m(1 -$
 237 $p)^{-\frac{1}{\alpha}}$, where $x_m > 0$, $\alpha > 0$, so $Q^{(\nu)}(p) \geq 0$ according to the
 238 chain rule.

239 **Theorem .5.** *A right-skewed continuous distribution with a*
 240 *monotonic decreasing pdf is second γ -ordered.*

241 *Proof.* A monotonic decreasing pdf means $f'(x) = F^{(2)}(x) \leq$
 242 0 . Since $Q'(p) \geq 0$, let $x = Q(F(x))$, then by differentiat-
 243 ing both sides of the equation twice, one can obtain $0 =$
 244 $Q^{(2)}(F(x))(F'(x))^2 + Q'(F(x))F^{(2)}(x) \Leftrightarrow Q^{(2)}(F(x)) =$
 245 $-\frac{Q'(F(x))F^{(2)}(x)}{(F'(x))^2} \geq 0$. The desired result is derived from Theo-
 246 rem .1 and .4. \square

247 Theorem .5 gives a valuable insight into the relation be-
 248 tween modality and orderliness. The conventional definition
 249 states that a distribution with a monotonic pdf is still con-
 250 sidered unimodal. However, within its supported interval, its
 251 mode number is zero. The number of modes and their magni-
 252 tudes are closely related to the possibility of orderliness being
 253 valid, although counterexamples can always be constructed
 254 for non-monotonic distributions. A proof of the second γ -
 255 orderliness, if $\gamma > 0$, can be easily established for the gamma
 256 distributions when $\alpha \leq 1$ as the pdf of the gamma distribution
 257 is $f(x) = \frac{\lambda^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\lambda}}}{\Gamma(\alpha)}$, where $x \geq 0$, $\lambda > 0$, $\alpha > 0$, Γ is the
 258 gamma function, it is a product of two monotonic decreasing
 259 functions under constraints. For $\alpha > 1$, the proof is challeng-
 260 ing, numerical results show that the orderliness is valid until
 261 $\alpha > 140$, the second orderliness is valid until $\alpha > 78$ (SI Text),
 262 but it is instructive to consider that when $\alpha \rightarrow \infty$ the gamma
 263 distribution converges to a Gaussian distribution with mean
 264 $\mu = \alpha\lambda$ and variance $\sigma = \alpha\lambda^2$.

265 **Theorem .6.** *If transforming a symmetric random variable*
 266 *X with a function $\phi(x)$ such that $\phi^{(2)}(x) \geq 0 \wedge \phi'(x) \geq 0$ over*
 267 *the interval supported, then the convex transformed distribution*
 268 *is ordered. If the quantile function of X satisfies $Q^{(2)}(\epsilon) \leq 0$,*
 269 *the convex transformed distribution is second ordered.*

270 *Proof.* Let $\phi \text{SQA}(\epsilon) = \frac{1}{2}(\phi(Q(\epsilon)) + \phi(Q(1 -$
 271 $\epsilon)))$, $0 \leq \epsilon \leq \frac{1}{2}$, then, $\frac{d\phi \text{SQA}}{d\epsilon} =$
 272 $\frac{1}{2}(\phi'(Q(\epsilon))Q'(\epsilon) - \phi'(Q(1 - \epsilon))Q'(1 - \epsilon))$
 273 $\frac{1}{2}Q'(\epsilon)(\phi'(Q(\epsilon)) - \phi'(Q(1 - \epsilon))) \leq 0$, since for a
 274 symmetric distribution, $m - Q(\epsilon) = Q(1 - \epsilon) - m$,
 275 differentiating both sides, $-Q'(\epsilon) = -Q'(1 - \epsilon)$,
 276 $Q'(\epsilon) \geq 0$, $\phi^{(2)}(x) \geq 0$. If further differentiating the
 277 equality, $Q^{(2)}(\epsilon) = -Q^{(2)}(1 - \epsilon)$. Since $\frac{d^{(2)}\phi \text{SQA}}{d\epsilon^{(2)}} =$
 278 $\frac{1}{2}(\phi^{(2)}(Q(\epsilon))(Q'(\epsilon))^2 + \phi^{(2)}(Q(1 - \epsilon))(Q'(1 - \epsilon))^2)$
 279 $\frac{1}{2}(\phi'(Q(\epsilon))(Q^2(\epsilon)) + \phi'(Q(1 - \epsilon))(Q^2(1 - \epsilon)))$
 280 $\frac{1}{2}((\phi^{(2)}(Q(\epsilon)) + \phi^{(2)}(Q(1 - \epsilon)))(Q'(\epsilon))^2)$
 281 $\frac{1}{2}((\phi'(Q(\epsilon)) - \phi'(Q(1 - \epsilon)))(Q^{(2)}(\epsilon)))$. If $Q^{(2)}(\epsilon) \leq 0$,
 282 $\frac{d^{(2)}\phi \text{SQA}}{d\epsilon^{(2)}} \geq 0$. \square

283 The mean-median-mode inequality for distributions of the
 284 powers and roots of the variates of a given distribution was
 285 investigated by Henry Rietz in 1927 (38), but the most trivial
 286 solution is the exponential transformation since the deriva-
 287 tives are always positive. An application of Theorem .6
 288 is that the lognormal distribution is ordered as it is expo-
 289 nentially transformed from the Gaussian distribution whose
 290 $Q^{(2)}(\epsilon) = -2\sqrt{2}\pi\sigma e^{2\text{erfc}^{-1}(2\epsilon)^2} \text{erfc}^{-1}(2\epsilon) \leq 0$, where erfc is
 291 the complementary error function (so, it is also second or-
 292 dered).

293 Theorem .6 also reveals a relation between convex transfor-
 294 mation and orderliness, since ϕ is the non-decreasing convex
 295 function in van Zwet's trailblazing work *Convex transforma-*
 296 *tions of random variables* (39). Consider a near-symmetric
 297 distribution S , such that SQA_{ϵ} as a function of ϵ fluctuates
 298 from 0 to $\frac{1}{2}$, with $\mu = m$. By definition, S is not ordered. Let
 299 s be the pdf of S . Applying the transformation $\phi(x)$ to S
 300 decreases $s(Q_S(\epsilon))$, and the decrease rate, due to the order,
 301 is much smaller for $s(Q_S(1 - \epsilon))$. As a consequence, as the
 302 second derivative of $\phi(x)$ increases, eventually, after a point,
 303 $s(Q_S(\epsilon))$ becomes greater than $s(Q_S(1 - \epsilon))$ even if it was not
 304 previously. Thus, the SQA_{ϵ} function becomes monotonically
 305 decreasing, and S becomes ordered. Accordingly, in a family
 306 of distributions that differ by a skewness-increasing transfor-
 307 mation in van Zwet's sense, violations of orderliness typically
 308 occur only when the distribution is near-symmetric.

309 Pearson, in 1895, proposed using the mean-median differ-
 310 ence $\mu - m$ as a measure of skewness after standardization
 311 (40). Bowley (1926) proposed a measure of skewness based on
 312 the SQA -median difference $\text{SQA}_{\epsilon} - m$ (41). Groeneveld and
 313 Meeden (1984) (42) generalized these measures of skewness
 314 based on van Zwet's convex transformation (39) and investi-
 315 gated their properties. A distribution is called monotonically
 316 right-skewed if and only if $\forall 0 \leq \epsilon_1 \leq \epsilon_2 \leq \frac{1}{2}, \text{SQA}_{\epsilon_1} - m \geq$
 317 $\text{SQA}_{\epsilon_2} - m$. Since m is a constant, the monotonic skewness is
 318 equivalent to the orderliness. The validity of robust measures
 319 of skewness based on the SQA -median difference is closely
 320 related to the orderliness of the distribution, because for a
 321 non-ordered distribution, the signs of $\text{SQA}_{\epsilon} - m$ with different
 322 breakdown points might be different. However, when the signs
 323 are different, that means for some measures of skewness, the
 324 distribution is left-skewed, others are right-skewed, therefore,
 325 the distribution is likely be generally near-symmetric, although
 326 counterexamples can be constructed.

327 **Data Availability.** Data for Figure ?? are given in SI Dataset
 328 S1. All codes have been deposited in [GitHub](#).

329 **ACKNOWLEDGMENTS.** I sincerely acknowledge the insightful
 330 comments from the editor which considerably elevated the lucidity
 331 and merit of this paper.

- 332 1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae*. (Henricus
 333 Dieterich), (1823).
- 334 2. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under
 335 partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
- 336 3. P Daniell, Observations weighted according to order. *Am. J. Math.* **42**, 222–236 (1920).
- 337 4. JW Tukey, A survey of sampling from contaminated distributions in *Contributions to probability*
 338 *and statistics*. (Stanford University Press), pp. 448–485 (1960).
- 339 5. WJ Dixon, Simplified Estimation from Censored Normal Samples. *The Annals Math. Stat.* **31**,
 340 385 – 391 (1960).
- 341 6. M Bieniek, Comparison of the bias of trimmed and winsorized means. *Commun. Stat. Methods*
 342 **45**, 6641–6650 (2016).
- 343 7. K Danielak, T Rychlik, Theory & methods: Exact bounds for the bias of trimmed means. *Aust.*
 344 *& New Zealand J. Stat.* **45**, 83–96 (2003).
- 345 8. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.*
 346 **34**, 598–611 (1963).

9. F Wilcoxon, Individual comparisons by ranking methods. *Biom. Bull.* 1, 80–83 (1945).
10. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* 35, 73–101 (1964).
11. Q Sun, WX Zhou, J Fan, Adaptive huber regression. *J. Am. Stat. Assoc.* 115, 254–265 (2020).
12. T Mathieu, Concentration study of m-estimators using the influence function. *Electron. J. Stat.* 16, 3695–3750 (2022).
13. AS Nemirovskij, DB Yudin, *Problem complexity and method efficiency in optimization.* (Wiley-Interscience), (1983).
14. MR Jerrum, LG Valiant, VV Vazirani, Random generation of combinatorial structures from a uniform distribution. *Theor. computer science* 43, 169–188 (1986).
15. N Alon, Y Matias, M Szegedy, The space complexity of approximating the frequency moments in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing.* pp. 20–29 (1996).
16. PL Bühlmann, Bagging, subagging and bragging for improving some prediction algorithms in *Research report/Seminar für Statistik, Eidgenössische Technische Hochschule (ETH).* (Seminar für Statistik, Eidgenössische Technische Hochschule (ETH), Zürich), Vol. 113, (2003).
17. JY Audibert, O Catoni, Robust linear least squares regression. *The Annals Stat.* 39, 2766–2794 (2011).
18. D Hsu, S Sabato, Heavy-tailed regression with a generalized median-of-means in *International Conference on Machine Learning.* (PMLR), pp. 37–45 (2014).
19. S Minsker, Geometric median and robust estimation in banach spaces. *Bernoulli* 21, 2308–2335 (2015).
20. C Brownlees, E Joly, G Lugosi, Empirical risk minimization for heavy-tailed losses. *The Annals Stat.* 43, 2507–2536 (2015).
21. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.* 44, 2695–2725 (2016).
22. E Joly, G Lugosi, Robust estimation of u-statistics. *Stoch. Process. their Appl.* 126, 3760–3773 (2016).
23. P Laforge, S Cléménçon, P Bertail, On medians of (randomized) pairwise means in *International Conference on Machine Learning.* (PMLR), pp. 1272–1281 (2019).
24. E Gobet, M Lerasle, D Métivier, Mean estimation for Randomized Quasi Monte Carlo method. working paper or preprint (2022).
25. B Efron, Bootstrap methods: Another look at the jackknife. *The Annals Stat.* 7, 1–26 (1979).
26. PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. *The annals statistics* 9, 1196–1217 (1981).
27. PJ Bickel, DA Freedman, Asymptotic normality and the bootstrap in stratified sampling. *The annals statistics* 12, 470–482 (1984).
28. R Helmers, P Janssen, N Veraverbeke, *Bootstrapping U-quantiles.* (CWI. Department of Operations Research, Statistics, and System Theory [BS]), (1990).
29. J Neyman, On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. Royal Stat. Soc.* 97, 558–606 (1934).
30. G McIntyre, A method for unbiased selective sampling, using ranked sets. *Aust. journal agricultural research* 3, 385–390 (1952).
31. C Stein, et al., Efficient nonparametric testing and estimation in *Proceedings of the third Berkeley symposium on mathematical statistics and probability.* Vol. 1, pp. 187–195 (1956).
32. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparametric models.* (Springer) Vol. 4, (1993).
33. 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).
37. KM Abadir, The mean-median-mode inequality: Counterexamples. *Econom. Theory* 21, 477–482 (2005).
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).
39. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
40. 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).
41. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
42. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* 33, 391–399 (1984).