Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on July 8, 2023

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

In 1823, Gauss (1) proved that for any unimodal distribution, $|m-\mu| \leq \sqrt{\frac{3}{4}}\omega$ and $\sigma \leq \omega \leq 2\sigma$, where μ is the population mean, m is the population median, ω is the root mean square deviation from the mode, and σ is the population standard deviation. This pioneering work revealed that, the potential bias of the median, the most fundamental robust location estimator, with respect to the mean is bounded in units of a scale parameter under certain assumptions. Bernard, Kazzi, and Vanduffel (2020) (2) further derived asymptotic bias bounds for any quantile in unimodal distributions with finite second moments. They showed that m has the smallest maximum distance to μ among all symmetric quantile averages (SQA_ϵ). Daniell, in 1920, (3) analyzed a class of estimators, linear combinations of order statistics, and identified that the ϵ-symmetric trimmed mean (STM_ϵ) belongs to this class. Another popular choice, the ϵ-symmetric Winsorized mean (SWM_ϵ), named after Winsor and introduced by Tukey (4) and Dixon (5) in 1960, is also an L-estimator. Bieniek (2016) derived exact bias upper bounds of the Winsorized mean based on Danielak and Rychlik's work (2003) on the trimmed mean for any distribution with a finite second moment and confirmed that the former is smaller than the latter (6, 7). In 1963, Hodges and Lehmann (8) proposed a class of nonparametric location estimators based on rank tests and, from the Wilcoxon signed-rank statistic (9), deduced the median of pairwise means as a robust location estimator for a symmetric population. Both L-statistics and R-statistics achieve robustness essentially by removing a certain proportion of extreme values. In 1964, Huber (10) generalized maximum likelihood estimation to the minimization of the sum of a specific loss function, which measures the residuals between the data points and the model's parameters. Some L-estimators are also M-estimators, e.g., the sample mean is an M-estimator with a squared error loss function, 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 the Huber M-estimator (11).

Quantile Average and Weighted Average

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

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

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

To extend the symmetric quantile average to the asymmetric case, two definitions for the ϵ,γ-quantile average ($\mathrm{QA}_{\epsilon,\gamma,n}$) are proposed. The first definition is:

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

and the second definition is:

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

where $\hat{Q}_n(p)$ is the empirical quantile function; γ is used to adjust the degree of asymmetry, $\gamma \geq 0$; and $0 \leq \epsilon \leq \frac{1}{1+\gamma}$. For trimming from both sides, [1] and [2] are essentially equivalent. The first definition along with $\gamma \geq 0$ and $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ are assumed in the rest of this article unless otherwise specified, since many common asymmetric distributions are right-skewed, and [1] allows trimming only from the right side by setting $\gamma = 0$.

Analogously, the weighted average can be defined as

$$\mathrm{WA}_{\epsilon,\gamma,n} := \frac{\int_0^{\frac{1}{1+\gamma}} \mathrm{QA}\left(\epsilon_0, \gamma, n\right) w(\epsilon_0) d\epsilon_0}{\int_0^{\frac{1}{1+\gamma}} w(\epsilon_0) d\epsilon_0}.$$

For any weighted average, if γ is omitted, it is assumed to be 1. The ϵ,γ-trimmed mean ($\mathrm{TM}_{\epsilon,\gamma,n}$) is a weighted average with a left trim size of $n\gamma\epsilon$ and a right trim size of $n\epsilon$, where $w(\epsilon_0) = \begin{cases} 0, & \epsilon_0 < \epsilon \\ 1, & \epsilon_0 \geq \epsilon \end{cases}$. Using this definition, regardless of whether $n\gamma\epsilon \notin \mathbb{N}$ or $n\epsilon \notin \mathbb{N}$, the TM computation remains the same, since this definition is based on the empirical quantile function. However, in this article, considering the computational cost in practice, non-asymptotic definitions of various types of weighted averages are primarily based on order statistics. Unless stated otherwise, the solution to their decimal issue is the same as that in SM.

Furthermore, for weighted averages, separating the breakdown point into upper and lower parts is necessary.

Definition .1 (Upper/lower breakdown point). The upper breakdown point is the breakdown point generalized in Davies and Gather (2005)'s paper (?). The finite-sample upper breakdown point is the finite sample breakdown point defined by Donoho and Huber (1983) (12) and also detailed in (?). The (finite-sample) lower breakdown point is replacing the infinity symbol in these definitions with negative infinity.

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

Classifying Distributions by the Signs of Derivatives

Let $\mathcal{P}_{\mathbb{R}}$ denote the set of all continuous distributions over $\mathbb{R}$ and $\mathcal{P}_{\mathbb{X}}$ denote the set of all discrete distributions over a countable set $\mathbb{X}$. The default of this article will be on the class of continuous distributions, $\mathcal{P}_{\mathbb{R}}$. However, it's worth noting that most discussions and results can be extended to encompass the discrete case, $\mathcal{P}_{\mathbb{X}}$, unless explicitly specified otherwise. Besides fully and smoothly parameterizing them by a Euclidean parameter or merely assuming regularity conditions, there exist additional methods for classifying distributions based on their characteristics, such as their skewness, peakedness, modality, and supported interval. In 1956, Stein initiated the study of estimating parameters in the presence of an infinite-dimensional nuisance shape parameter (13) and proposed a necessary condition for this type of problem, a contribution later explicitly recognized as initiating the field of semiparametric statistics (14). In 1982, Bickel simplified Stein's general heuristic necessary condition (13), derived sufficient conditions, and used them in formulating adaptive estimates (14). A notable example discussed in these groundbreaking works was the adaptive estimation of the center of symmetry for an unknown symmetric distribution, which is a semiparametric model. In 1993, Bickel, Klaassen, Ritov, and Wellner published an influential semiparametrics textbook (15), which categorized most common statistical models as semiparametric models, considering parametric and nonparametric models as two special cases within this classification. Yet, 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., $(f'(x) > 0 \text{ for } x \leq M) \wedge (f'(x) < 0 \text{ for } x \geq M)$, where $f(x)$ is the probability density function (pdf) of a random variable X, M is the mode. 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) (16, 17) endeavored to determine sufficient conditions for the mean-median-mode inequality to hold, thereby implying the possibility of its violation. The class of unimodal distributions that satisfy the mean-median-mode inequality constitutes a subclass of $\mathcal{P}_U$, denoted by $\mathcal{P}_{MMM} \subsetneq \mathcal{P}_U$. To further investigate the relations of location estimates within a distribution, the γ-orderliness for a right-skewed distribution is defined as

$$\forall 0 \leq \epsilon_1 \leq \epsilon_2 \leq \frac{1}{1+\gamma}, \mathrm{QA}(\epsilon_1, \gamma) \geq \mathrm{QA}(\epsilon_2, \gamma).$$

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

Theorem .1. *A distribution is γ-ordered if and only if its 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 $0 \leq \epsilon \leq \frac{1}{1+\gamma}$.*

Proof. Without loss of generality, consider the case of right-skewed distribution. From the above definition of γ-orderliness, it is deduced that $\frac{Q(\gamma\epsilon-\delta)+Q(1-\epsilon+\delta)}{2} \geq \frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2} \Leftrightarrow Q(\gamma\epsilon-\delta) - Q(\gamma\epsilon) \geq Q(1-\epsilon) - Q(1-\epsilon+\delta) \Leftrightarrow Q'(1-\epsilon) \geq Q'(\gamma\epsilon)$, where δ is an infinitesimal positive quantity. Observing that the quantile function is the inverse function of the cumulative distribution function (cdf), $Q'(1-\epsilon) \geq Q'(\gamma\epsilon) \Leftrightarrow F'(Q(\gamma\epsilon)) \geq F'(Q(1-\epsilon))$, thereby completing the proof, since the derivative of cdf is pdf. □

According to Theorem .1, if a probability distribution is right-skewed and monotonic decreasing, it will always be γ-ordered. For a right-skewed unimodal distribution, if $Q(\gamma\epsilon) > M$, then the inequality $f(Q(\gamma\epsilon)) \geq f(Q(1-\epsilon))$ holds. The principle is extendable to unimodal-like distributions. Suppose there is a right-skewed unimodal-like distribution with the first mode, denoted as M_1, having the greatest probability density, while there are several smaller modes located towards the higher values of the distribution. Furthermore, assume that this distribution follows the mean-γ-median-first mode inequality, amd 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 $Q(\gamma\epsilon) > M_1$, the inequality $f(Q(\gamma\epsilon)) \geq f(Q(1-\epsilon))$ also holds. In other words, even though a distribution following the mean-γ-median-mode inequality may not be strictly γ-ordered, the inequality defining the γ-orderliness remains valid for most quantile averages. The mean-γ-median-mode inequality can also indicate possible bounds for γ in practice, e.g., for any distributions, when $\gamma \to \infty$, the γ-median will be greater than the mean and the mode, when $\gamma \to 0$, the γ-median will be smaller than the mean and the mode, a reasonable γ should maintain the validity of the mean-γ-median-mode inequality.

The definition above of γ-orderliness for a right-skewed distribution implies a monotonic decreasing behavior of the quantile average function with respect to the breakdown point. Therefore, consider the sign of the partial derivative, it can also be expressed as:

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

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

Furthermore, many common right-skewed distributions, such as the Weibull, gamma, lognormal, and Pareto distributions, are partially bounded, indicating a convex behavior of the QA function with respect to ϵ as ϵ approaches 0. By further assuming convexity, the second γ-orderliness can be defined for a right-skewed distribution as follows,

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

Analogously, the νth γ-orderliness of a right-skewed distribution can be defined as $(-1)^{\nu} \frac{\partial^{\nu} \mathrm{QA}}{\partial \epsilon^{\nu}} \geq 0 \wedge \ldots \wedge -\frac{\partial \mathrm{QA}}{\partial \epsilon} \geq 0$. If $\gamma = 1$, the νth γ-orderliness is referred as to νth orderliness. Let $\mathcal{P}_O$ denote the set of all distributions that are ordered and $\mathcal{P}_{O_\nu}$ and $\mathcal{P}_{\gamma O_\nu}$ represent the sets of all distributions that are νth ordered and νth γ-ordered, respectively. When the shape parameter of the Weibull distribution, α, is smaller than 3.258, it can be shown that the Weibull distribution belongs to $\mathcal{P}_U \cap \mathcal{P}_O \cap \mathcal{P}_{O_2} \cap \mathcal{P}_{O_3}$ (SI Text). At $\alpha \approx 3.602$, the Weibull distribution is symmetric, and as $\alpha \to \infty$, the skewness of the Weibull distribution approaches 1. Therefore, the parameters that prevent it from being 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 intricate and, at times, challenging. The following theorems establish the relationship between $\mathcal{P}_O$, $\mathcal{P}_{O_\nu}$, and $\mathcal{P}_{\gamma O_\nu}$, and a wide range of other semi-parametric distributions. They can be used to quickly identify some parametric distributions in $\mathcal{P}_O$, $\mathcal{P}_{O_\nu}$, 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 νth γ-ordered if and only if the family of probability distributions is νth γ-ordered.*

Proof. Let Q_0 denote the quantile function of the standard distribution without any shifts or scaling. After a location-scale transformation, the quantile function becomes $Q(p) = \lambda Q_0(p)+\mu$, where λ is the scale parameter and μ is the location parameter. According to the definition of the νth γ-orderliness, the signs of derivatives of the QA function are invariant after this transformation. As the location-scale transformation is reversible, the proof is complete. $\square$

Theorem .2 demonstrates that in the analytical proof of the νth 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.

Theorem .3. *Define a γ-symmetric distribution as one for which the quantile function satisfies $Q(\gamma\epsilon) = 2Q(\frac{\gamma}{1+\gamma}) - Q(1-\epsilon)$ for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$. Any γ-symmetric distribution is νth γ-ordered.*

Proof. The equality, $Q(\gamma\epsilon) = 2Q(\frac{\gamma}{1+\gamma}) - Q(1-\epsilon)$, implies that $\frac{\partial Q(\gamma\epsilon)}{\partial\epsilon} = \gamma Q'(\gamma\epsilon) = \frac{\partial(-Q(1-\epsilon))}{\partial\epsilon} = Q'(1-\epsilon)$. From the first definition of QA, the QA function of the γ-symmetric distribution is a horizontal line, since $\frac{\partial \mathrm{QA}}{\partial\epsilon} = \gamma Q'(\gamma\epsilon) - Q'(1-\epsilon) = 0$. So, the νth order derivative of QA is always zero. $\square$

Theorem .4. *A symmetric distribution is a special case of the γ-symmetric distribution when $\gamma = 1$, provided that the cdf is monotonic.*

Proof. A symmetric distribution is a probability distribution such that for all x, $f(x) = f(2m-x)$. Its cdf satisfies $F(x) = 1 - F(2m-x)$. Let $x = Q(p)$, then, $F(Q(p)) = p = 1 - F(2m - Q(p))$ and $F(Q(1-p)) = 1-p \Leftrightarrow p = 1 - F(Q(1-p))$. Therefore, $F(2m - Q(p)) = F(Q(1-p))$. Since the cdf is monotonic, $2m - Q(p) = Q(1-p) \Leftrightarrow Q(p) = 2m - Q(1-p)$. Choosing $p = \epsilon$ yields the desired result. $\square$

Since the generalized Gaussian distribution is symmetric around the median, it is νth ordered, as a consequence of Theorem .3.

Theorem .5. *Any right-skewed distribution whose quantile function Q satisfies $Q^{(\nu)}(p) \geq 0 \wedge \ldots Q^{(i)}(p) \geq 0 \ldots \wedge Q^{(2)}(p) \geq 0$, i mod $2 = 0$, is νth γ-ordered, provided that $0 \leq \gamma \leq 1$.*

Proof. Since $(-1)^i \frac{\partial^i \mathrm{QA}}{\partial\epsilon^i} = \frac{1}{2}((-\gamma)^i Q^i(\gamma\epsilon) + Q^i(1-\epsilon))$ and $1 \leq i \leq \nu$, when i mod $2 = 0$, $(-1)^i \frac{\partial^i \mathrm{QA}}{\partial\epsilon^i} \geq 0$ for all $\gamma \geq 0$. When i mod $2 = 1$, if further assuming $0 \leq \gamma \leq 1$, $(-1)^i \frac{\partial^i \mathrm{QA}}{\partial\epsilon^i} \geq 0$, since $Q^{(i+1)}(p) \geq 0$. $\square$

This result makes it straightforward to show that the Pareto distribution follows the νth γ-orderliness, provided that $0 \leq \gamma \leq 1$, since the quantile function of the Pareto distribution is $Q(p) = x_m(1-p)^{-\frac{1}{\alpha}}$, where $x_m > 0$, $\alpha > 0$, and so $Q^{(\nu)}(p) \geq 0$ for all $\nu \in \mathbb{N}$ according to the chain rule.

Theorem .6. *A right-skewed distribution with a monotonic decreasing pdf is second γ-ordered.*

Proof. Given that a monotonic decreasing pdf implies $f'(x) = F^{(2)}(x) \leq 0$, let $x = Q(F(x))$, then by differentiating both sides of the equation twice, one can obtain $0 = Q^{(2)}(F(x))(F'(x))^2 + Q'(F(x))F^{(2)}(x) \Rightarrow Q^{(2)}(F(x)) = -\frac{Q'(F(x))F^{(2)}(x)}{(F'(x))^2} \geq 0$, since $Q'(p) \geq 0$. Theorem .1 already established the γ-orderliness for all $\gamma \geq 0$, which means $\forall 0 \leq \epsilon \leq \frac{1}{1+\gamma}$, $\frac{\partial \mathrm{QA}}{\partial\epsilon} \leq 0$. The desired result is then derived from the proof of Theorem .5, since $(-1)^2 \frac{\partial^2 \mathrm{QA}}{\partial\epsilon^2} \geq 0$ for all $\gamma \geq 0$. $\square$

Theorem .6 provides valuable insights into the relation between modality and second γ-orderliness. The conventional definition states that a distribution with a monotonic pdf is still considered unimodal. However, within its supported interval, the mode number is zero. Theorem .1 implies that the number of modes and their magnitudes within a distribution are closely related to the likelihood of γ-orderliness being valid. This is because, for a distribution satisfying the necessary and sufficient condition in Theorem .1, it is already implied that the probability density of the left-hand side of the γ-median is always greater than the corresponding probability density of the right-hand side of the γ-median, so although counterexamples can always be constructed for non-monotonic distributions, the general shape of a γ-ordered distribution should have a single dominant mode. It can be easily established that the gamma distribution is second γ-ordered when $\alpha \leq 1$, as the pdf of the gamma distribution is $f(x) = \frac{\lambda^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\lambda}}}{\Gamma(\alpha)}$, where $x \geq 0$, $\lambda > 0$, $\alpha > 0$, and Γ represents the gamma function. This pdf is a product of two monotonic decreasing functions under constraints. For $\alpha > 1$, analytical analysis becomes challenging. Numerical results show that orderliness is valid until $\alpha > 00.000$, the second orderliness is valid until $\alpha > 00.000$, and the third orderliness is valid until $\alpha > 00.000$ (SI Text). 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 $\sigma = \alpha\lambda^2$. The skewness of the gamma distribution, $\frac{\alpha+2}{\sqrt{\alpha(\alpha+1)}}$, is monotonic with respect to α, since $\frac{\partial\tilde{\mu}_3(\alpha)}{\partial\alpha} = \frac{-3\alpha-2}{2(\alpha(\alpha+1))^{3/2}} < 0$. When $\alpha = 00.000$, $\tilde{\mu}_3(\alpha) = 1.027$. Theorefore, similar to the Weibull distribution, the parameters which make these distributions fail to be included in $\mathcal{P}_U \cap \mathcal{P}_O \cap \mathcal{P}_{O_2} \cap \mathcal{P}_{O_3}$ also correspond to cases when it is near-symmetric.

Theorem .7. *Consider a γ-symmetric random variable X. Let it be transformed using a function $\phi(x)$ such that $\phi^{(2)}(x) \geq 0$ over the interval supported, the resulting convex transformed distribution is γ-ordered. Moreover, if the quantile function of X satisfies $Q^{(2)}(p) \leq 0$, the convex transformed distribution is second γ-ordered.*

Proof. Let $\phi\mathrm{QA}(\epsilon,\gamma) = \frac{1}{2}(\phi(Q(\gamma\epsilon)) + \phi(Q(1-\epsilon)))$. Then, for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$, $\frac{\partial\phi\mathrm{QA}}{\partial\epsilon} =$

$\frac{1}{2}\left(\gamma\phi'\left(Q\left(\gamma\epsilon\right)\right)Q'\left(\gamma\epsilon\right)-\phi'\left(Q\left(1-\epsilon\right)\right)Q'\left(1-\epsilon\right)\right) = \frac{1}{2}\gamma Q'\left(\gamma\epsilon\right)\left(\phi'\left(Q\left(\gamma\epsilon\right)\right)-\phi'\left(Q\left(1-\epsilon\right)\right)\right) \leq 0$, since for a γ-symmetric distribution, $Q(\frac{1}{1+\gamma})-Q\left(\gamma\epsilon\right)=Q\left(1-\epsilon\right)-Q(\frac{1}{1+\gamma})$, differentiating both sides, $-\gamma Q'\left(\gamma\epsilon\right) = -Q'(1-\epsilon)$, where $Q'\left(p\right) \geq 0, \phi^{(2)}\left(x\right) \geq 0$. If further differentiating the equality, $\gamma^2 Q^{(2)}\left(\gamma\epsilon\right) = -Q^{(2)}(1-\epsilon)$. Since $\frac{\partial^{(2)}\phi\mathrm{QA}}{\partial\epsilon^{(2)}} = \frac{1}{2}\left(\gamma^2\phi^2\left(Q\left(\gamma\epsilon\right)\right)\left(Q'\left(\gamma\epsilon\right)\right)^2+\phi^2\left(Q\left(1-\epsilon\right)\right)\left(Q'\left(1-\epsilon\right)\right)^2\right) + \frac{1}{2}\left(\gamma^2\phi'\left(Q\left(\gamma\epsilon\right)\right)\left(Q^2\left(\gamma\epsilon\right)\right)+\phi'\left(Q\left(1-\epsilon\right)\right)\left(Q^2\left(1-\epsilon\right)\right)\right) = \frac{1}{2}\left(\left(\phi^{(2)}\left(Q\left(\gamma\epsilon\right)\right)+\phi^{(2)}\left(Q\left(1-\epsilon\right)\right)\right)\left(\gamma^2 Q'\left(\gamma\epsilon\right)\right)^2\right) + \frac{1}{2}\left(\left(\phi'\left(Q\left(\gamma\epsilon\right)\right)-\phi'\left(Q\left(1-\epsilon\right)\right)\right)\gamma^2 Q^{(2)}\left(\gamma\epsilon\right)\right)$. If $Q^{(2)}\left(p\right)\leq 0$, for all $0\leq\epsilon\leq\frac{1}{1+\gamma}$, $\frac{\partial^{(2)}\phi\mathrm{QA}}{\partial\epsilon^{(2)}}\geq 0$. □

An application of Theorem .7 is that the lognormal distribution is ordered as it is exponentially transformed from the Gaussian distribution. The quantile function of the Gaussian distribution meets the condition $Q^{(2)}\left(p\right) = -2\sqrt{2}\pi\sigma e^{2\mathrm{erfc}^{-1}(2p)^2}\mathrm{erfc}^{-1}(2p)\leq 0$, where σ is the standard deviation of the Gaussian distribution and erfc denotes the complementary error function. Thus, the lognormal distribution is second ordered. Numerical results suggest that it is also third ordered, although analytically proving this result is challenging.

Theorem .7 also reveals a relation between convex transformation and orderliness, since ϕ is the non-decreasing convex function in van Zwet's trailblazing work *Convex transformations of random variables* (18) if adding an additional constraint that $\phi'\left(x\right)\geq 0$. Consider a near-symmetric distribution S, such that the $\mathrm{SQA}(\epsilon)$ as a function of ϵ fluctuates from 0 to $\frac{1}{2}$. By definition, S is not ordered. Let s be the pdf of S. Applying the transformation $\phi\left(x\right)$ to S decreases $s(Q_S(\epsilon))$, and the decrease rate, due to the order, is much smaller for $s(Q_S(1-\epsilon))$. As a consequence, as $\phi^{(2)}\left(x\right)$ increases, eventually, after a point, for all $0\leq\epsilon\leq\frac{1}{1+\gamma}$, $s(Q_S(\epsilon))$ becomes greater than $s(Q_S(1-\epsilon))$ even if it was not previously. Thus, the $\mathrm{SQA}(\epsilon)$ function becomes monotonically decreasing, and S becomes ordered. Accordingly, in a family of distributions that differ by a skewness-increasing transformation in van Zwet's sense, violations of orderliness typically occur only when the distribution is near-symmetric.

Pearson proposed using the 3 times standardized mean-median difference, $\frac{3(\mu-m)}{\sigma}$, as a measure of skewness in 1895 (19). Bowley (1926) proposed a measure of skewness based on the $\mathrm{SQA}_{\epsilon=\frac{1}{4}}$-median difference $\mathrm{SQA}_{\epsilon=\frac{1}{4}}-m$ (20). Groeneveld and Meeden (1984) (21) generalized these measures of skewness based on van Zwet's convex transformation (18) while exploring their properties. A distribution is called monotonically right-skewed if and only if $\forall 0\leq\epsilon_1\leq\epsilon_2\leq\frac{1}{2}, \mathrm{SQA}_{\epsilon_1}-m\geq \mathrm{SQA}_{\epsilon_2}-m$. Since m is a constant, the monotonic skewness is equivalent to the orderliness. For a nonordered distribution, the signs of $\mathrm{SQA}_{\epsilon}-m$ with different breakdown points might be different, implying that some skewness measures indicate left-skewed distribution, while others suggest right-skewed distribution. Although it seems reasonable that such a distribution is likely be generally near-symmetric, counterexamples can be constructed. For example, first consider the Weibull distribution, when $\alpha>\frac{1}{1-\ln(2)}$, it is near-symmetric and nonordered, the non-monotonicity of the SQA function arises when ϵ is close to $\frac{1}{2}$, but if then replacing the third quartile with one from a right-skewed heavy-tailed distribution leads to a right-skewed, heavy-tailed, and nonordered distribution. Therefore, the validity of robust measures of skewness based on the SQA-median difference is closely related to the orderliness of the distribution.

Remarkably, in 2018, Li, Shao, Wang, Yang (22) proved the bias bound of any quantile for arbitrary continuous distributions with finite second moments. Here, let $\mathcal{P}_{\mu,\sigma}$ denotes the set of continuous distributions whose mean is μ and standard deviation is σ. The bias upper bound of the quantile average for $P\in\mathcal{P}_{\mu=0,\sigma=1}$ is given in the following theorem.

Theorem .8. *The bias upper bound of the quantile average for any continuous distribution whose mean is zero and standard deviation is one is*

$$\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}} QA(\epsilon,\gamma)=\frac{1}{2}\left(\sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}}+\sqrt{\frac{1-\epsilon}{\epsilon}}\right),$$

where $0\leq\epsilon\leq\frac{1}{1+\gamma}$.

Proof. Since $\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}}\frac{1}{2}(Q(\gamma\epsilon)+Q(1-\epsilon)) \leq \frac{1}{2}(\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}}Q(\gamma\epsilon) + \sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}}Q(1-\epsilon))$, the assertion follows directly from the Lemma 2.6 in (22). □

In 2020, Bernard et al. (2) further refined these bounds for unimodal distributions and derived the bias bound of the symmetric quantile average. Here, the bias upper bound of the quantile average, $0\leq\gamma<5$, for $P\in\mathcal{P}_U\cap\mathcal{P}_{\mu=0,\sigma=1}$ is given as

$$\sup_{P\in\mathcal{P}_U\cap\mathcal{P}_{\mu=0,\sigma=1}}\mathrm{QA}(\epsilon,\gamma)=\begin{cases}\frac{1}{2}\left(\sqrt{\frac{4}{9\epsilon}-1}+\sqrt{\frac{3\gamma\epsilon}{4-3\gamma\epsilon}}\right) & 0\leq\epsilon\leq\frac{1}{6}\\ \frac{1}{2}\left(\sqrt{\frac{3(1-\epsilon)}{4-3(1-\epsilon)}}+\sqrt{\frac{3\gamma\epsilon}{4-3\gamma\epsilon}}\right) & \frac{1}{6}<\epsilon\leq\frac{1}{1+\gamma}.\end{cases}$$

The proof based on the bias bounds of any quantile (2) and the $\gamma\geq 5$ case are given in the SI Text. Subsequent theorems reveal the safeguarding role these bounds play in defining estimators based on νth γ-orderliness. The proof of Theorem .9 is provided in the SI Text.

Theorem .9. $\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}} QA(\epsilon,\gamma)$ *is monotonic decreasing with respect to* ϵ *over* $[0,\frac{1}{1+\gamma}]$, *provided that* $0\leq\gamma\leq 1$.

Theorem .10. $\sup_{P\in\mathcal{P}_U\cap\mathcal{P}_{\mu=0,\sigma=1}} QA(\epsilon,\gamma)$ *is a nonincreasing function with respect to* ϵ *on the interval* $[0,\frac{1}{1+\gamma}]$, *provided that* $0\leq\gamma\leq 1$.

Proof. When $0\leq\epsilon\leq\frac{1}{6}$, $\frac{\partial\sup\mathrm{QA}}{\partial\epsilon} = \frac{\gamma}{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2} - \frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2} = \frac{\sqrt{\gamma}}{\sqrt{\frac{\epsilon}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2} - \frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2}$. If $\gamma=0$ and $\epsilon\to 0^+$, $\frac{\partial\sup\mathrm{QA}}{\partial\epsilon} = -\frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2} < 0$. If $\epsilon\to 0^+$, $\lim_{\epsilon\to 0^+}\left(\frac{\gamma}{(4-3\gamma\epsilon)^2\sqrt{\frac{\epsilon\gamma}{12-9\gamma\epsilon}}}-\frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2}\right) = \lim_{\epsilon\to 0^+}\left(\frac{\sqrt{3\gamma}}{\sqrt{4^3\epsilon}}-\frac{1}{6\sqrt{\epsilon^3}}\right) \to -\infty$, for all $0\leq\gamma\leq 1$, so, $\frac{\partial\sup\mathrm{QA}}{\partial\epsilon} < 0$. When $0<\epsilon\leq\frac{1}{6}$ and $0<\gamma\leq 1$, to prove $\frac{\partial\sup\mathrm{QA}}{\partial\epsilon}\leq 0$, it is equivalent to showing $\frac{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2}{\gamma} \geq 3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2$. Define $L(\epsilon,\gamma) = \frac{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2}{\gamma}$, $R(\epsilon,\gamma) = 3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2$. $\frac{L(\epsilon,\gamma)}{\epsilon^2} = \frac{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2}{\gamma\epsilon^2} = \frac{1}{\gamma}\left(\frac{4}{\epsilon}-3\gamma\right)^2\sqrt{\frac{1}{\frac{12}{\epsilon\gamma}-9}}$,

$\frac{R(\epsilon,\gamma)}{\epsilon^2} = 3\sqrt{\frac{4}{\epsilon} - 9}$. Then, the above inequality is equivalent to $\frac{L(\epsilon,\gamma)}{\epsilon^2} \geq \frac{R(\epsilon,\gamma)}{\epsilon^2} \Leftrightarrow \frac{1}{\gamma}\sqrt{\frac{1}{\frac{12}{\epsilon\gamma}-9}}\left(\frac{4}{\epsilon} - 3\gamma\right)^2 \geq 3\sqrt{\frac{4}{\epsilon} - 9} \Leftrightarrow \frac{1}{\gamma}\left(\frac{4}{\epsilon} - 3\gamma\right)^2 \geq 3\sqrt{\frac{12}{\epsilon\gamma} - 9}\sqrt{\frac{4}{\epsilon} - 9} \Leftrightarrow \frac{1}{\gamma^2}\left(\frac{4}{\epsilon} - 3\gamma\right)^4 \geq 9\left(\frac{12}{\epsilon\gamma} - 9\right)\left(\frac{4}{\epsilon} - 9\right)$. Let $LmR\left(\frac{1}{\epsilon}\right) = \frac{1}{\gamma^2}\left(\frac{4}{\epsilon} - 3\gamma\right)^4 - 9\left(\frac{12}{\epsilon\gamma} - 9\right)\left(\frac{4}{\epsilon} - 9\right)$. $\frac{\partial LmR(1/\epsilon)}{\partial(1/\epsilon)} = \frac{16\left(\frac{4}{\epsilon}-3\gamma\right)^3}{\gamma^2} - 36\left(\frac{12}{\epsilon\gamma} - 9\right) - \frac{108\left(4\frac{4}{\epsilon}-9\right)}{\gamma} = \frac{4\left(4(\frac{4}{\epsilon}-3\gamma)^3-27\gamma(\frac{4}{\epsilon}-3\gamma)+27(9-\frac{4}{\epsilon})\gamma\right)}{\gamma^2} = \frac{4\left(256\frac{1}{\epsilon}^3-576\frac{1}{\epsilon}^2\gamma+432\frac{1}{\epsilon}\gamma^2-216\frac{1}{\epsilon}\gamma-108\gamma^3+81\gamma^2+243\gamma\right)}{\gamma^2}$. Since $256\frac{1}{\epsilon}^3 - 576\frac{1}{\epsilon}^2\gamma + 432\frac{1}{\epsilon}\gamma^2 - 216\frac{1}{\epsilon}\gamma - 108\gamma^3 + 81\gamma^2 + 243\gamma \geq 1536\frac{1}{\epsilon}^2 - 576\frac{1}{\epsilon}^2 + 432\frac{1}{\epsilon}\gamma^2 - 216\frac{1}{\epsilon}\gamma - 108\gamma^3 + 81\gamma^2 + 243\gamma \geq 924\frac{1}{\epsilon}^2 + 36\frac{1}{\epsilon}^2 - 216\frac{1}{\epsilon} + 432\frac{1}{\epsilon}\gamma^2 - 108\gamma^3 + 81\gamma^2 + 243\gamma \geq 924\frac{1}{\epsilon}^2 + 36\frac{1}{\epsilon}^2 - 216\frac{1}{\epsilon} + 513\gamma^2 - 108\gamma^3 + 243\gamma > 0$, $\frac{\partial LmR(1/\epsilon)}{\partial(1/\epsilon)} > 0$. Also, $LmR(6) = \frac{81(\gamma-8)\left((\gamma-8)^3+15\gamma\right)}{\gamma^2} > 0 \Longleftrightarrow \gamma^4 - 32\gamma^3 + 399\gamma^2 - 2168\gamma + 4096 > 0$. If $0 < \gamma \leq 1$, then $32\gamma^3 < 256$. Also, $\gamma^4 > 0$. So, it suffices to prove that $399\gamma^2 - 2168\gamma + 4096 > 256$. Applying the quadratic formula demonstrates the validity of $LmR(6) > 0$, if $0 < \gamma \leq 1$. Hence, $LmR\left(\frac{1}{\epsilon}\right) \geq 0$ for $\epsilon \in (0, \frac{1}{6}]$, if $0 < \gamma \leq 1$. The first part is finished.

When $\frac{1}{6} < \epsilon \leq \frac{1}{1+\gamma}$, $\frac{\partial \sup \mathrm{QA}}{\partial \epsilon} = \sqrt{3}\left(\frac{\gamma}{\sqrt{\gamma\epsilon}(4-3\gamma\epsilon)^{\frac{3}{2}}} - \frac{1}{\sqrt{1-\epsilon}(3\epsilon+1)^{\frac{3}{2}}}\right)$. If $\gamma = 0$, $\frac{\gamma}{\sqrt{\gamma\epsilon}(4-3\gamma\epsilon)^{\frac{3}{2}}} = \frac{\sqrt{\gamma}}{\sqrt{\epsilon}(4-3\gamma\epsilon)^{\frac{3}{2}}} = 0$, so $\frac{\partial \sup \mathrm{QA}}{\partial \epsilon} = \sqrt{3}\left(-\frac{1}{\sqrt{1-\epsilon}(3\epsilon+1)^{\frac{3}{2}}}\right) < 0$, for all $\frac{1}{6} < \epsilon \leq \frac{1}{1+\gamma}$. If $\gamma > 0$, to determine whether $\frac{\partial \sup \mathrm{QA}}{\partial \epsilon} \leq 0$, when $\frac{1}{6} < \epsilon \leq \frac{1}{1+\gamma}$, since $\sqrt{1-\epsilon}\,(3\epsilon+1)^{\frac{3}{2}} > 0$ and $\sqrt{\gamma\epsilon}\,(4-3\gamma\epsilon)^{\frac{3}{2}} > 0$, showing $\frac{\sqrt{\gamma\epsilon}(4-3\gamma\epsilon)^{\frac{3}{2}}}{\gamma} \geq \sqrt{1-\epsilon}\,(3\epsilon+1)^{\frac{3}{2}} \Leftrightarrow \frac{\gamma\epsilon(4-3\gamma\epsilon)^3}{\gamma^2} \geq (1-\epsilon)(3\epsilon+1)^3 \Leftrightarrow -27\gamma^2\epsilon^4 + 108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} + 27\epsilon^4 - 162\epsilon^2 - 8\epsilon - 1 \geq 0$ is sufficient. When $0 < \gamma \leq 1$, the inequality can be further simplified to $108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} - 162\epsilon^2 - 8\epsilon - 1 \geq 0$. Since $\epsilon \leq \frac{1}{1+\gamma}$, $\gamma \leq \frac{1}{\epsilon} - 1$. Also, as $0 < \gamma \leq 1$ is assumed, the range of γ can be expressed as $0 < \gamma \leq \min(1, \frac{1}{\epsilon} - 1)$. When $\frac{1}{6} < \epsilon \leq \frac{1}{2}$, $1 < \frac{1}{\epsilon} - 1$, so in this case, $0 < \gamma \leq 1$. When $\frac{1}{2} \leq \epsilon < 1$, so in this case, $0 < \gamma \leq \frac{1}{\epsilon} - 1$. Let $h(\gamma) = 108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma}$, $\frac{\partial h(\gamma)}{\partial \gamma} = 108\epsilon^3 - \frac{64\epsilon}{\gamma^2}$. When $\gamma \leq \sqrt{\frac{64\epsilon}{18\epsilon^3}}$, $\frac{\partial h(\gamma)}{\partial \gamma} \geq 0$, when $\gamma \geq \sqrt{\frac{64\epsilon}{18\epsilon^3}}$, $\frac{\partial h(\gamma)}{\partial \gamma} \leq 0$, therefore, the minimum of $h(\gamma)$ must be when γ is equal to the boundary point of the domain. When $\frac{1}{6} < \epsilon \leq \frac{1}{2}$, $0 < \gamma \leq 1$, since $h(0) \to \infty$, $h(1) = 108\epsilon^3 + 64\epsilon$, the minimum occurs at the boundary point $\gamma = 1$, $108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} - 162\epsilon^2 - 8\epsilon - 1 > 108\epsilon^3 + 56\epsilon - 162\epsilon^2 - 1$. Let $g(\epsilon) = 108\epsilon^3 + 56\epsilon - 162\epsilon^2 - 1$. $g'(\epsilon) = 324\epsilon^2 - 324\epsilon + 56$, when $\epsilon \leq \frac{2}{9}$, $g'(\epsilon) \geq 0$, when $\frac{2}{9} \leq \epsilon \leq \frac{1}{2}$, $g'(\epsilon) \leq 0$, since $g(\frac{1}{6}) = \frac{13}{3}$, $g(\frac{1}{2}) = 0$, so $g(\epsilon) \geq 0$, $108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} - 162\epsilon^2 - 8\epsilon - 1 \geq 0$. When $\frac{1}{2} \leq \epsilon < 1$, $0 < \gamma \leq \frac{1}{\epsilon} - 1$. Since $h(\frac{1}{\epsilon} - 1) = 108(\frac{1}{\epsilon} - 1)\epsilon^3 + \frac{64\epsilon}{\frac{1}{\epsilon}-1}$, $108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} - 162\epsilon^2 - 8\epsilon - 1 > 108\left(\frac{1}{\epsilon} - 1\right)\epsilon^3 + \frac{64\epsilon}{\frac{1}{\epsilon}-1} - 162\epsilon^2 - 8\epsilon - 1 = \frac{-108\epsilon^4+54\epsilon^3-18\epsilon^2+7\epsilon+1}{\epsilon-1}$. Let $nu(\epsilon) = -108\epsilon^4 + 54\epsilon^3 - 18\epsilon^2 + 7\epsilon + 1$, then $nu'(\epsilon) = -432\epsilon^3 + 162\epsilon^2 - 36\epsilon + 7$, $nu''(\epsilon) = -1296\epsilon^2 + 324\epsilon - 36 < 0$. Since $nu'(\epsilon = \frac{1}{2}) = -\frac{49}{2} < 0$, $nu'(\epsilon) < 0$. Also, $nu(\epsilon = \frac{1}{2}) = 0$, so $nu(\epsilon) \geq 0$, $108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} - 162\epsilon^2 - 8\epsilon - 1 \geq 0$ is also valid. As a result, this simplified inequality is valid within the range of $\frac{1}{6} < \epsilon \leq \frac{1}{1+\gamma}$, when $0 < \gamma \leq 1$. Then, it validates $\frac{\partial \sup \mathrm{QA}}{\partial \epsilon} \leq 0$ for the same range of ϵ and γ.

The first and second formulae, when $\epsilon = \frac{1}{6}$, are all equal to $\frac{1}{2}\left(\frac{\sqrt{\frac{\gamma}{4-\frac{\gamma}{2}}}}{\sqrt{2}} + \sqrt{\frac{5}{3}}\right)$. It follows that $\sup \mathrm{QA}(\epsilon, \gamma)$ is continuous over $[0, \frac{1}{1+\gamma}]$. Hence, $\frac{\partial \sup \mathrm{QA}}{\partial \epsilon} \leq 0$ holds for the entire range $0 \leq \epsilon \leq \frac{1}{1+\gamma}$, when $0 \leq \gamma \leq 1$, which leads to the assertion of this theorem. $\square$

Let $\mathcal{P}_{\Upsilon}^{k}$ denote the set of all continuous distributions whose moments, from the first to the kth, are all finite. For a right-skewed distribution, it suffices to consider the upper bound. The monotonicity of $\sup_{P\in\mathcal{P}_{\Upsilon}^{2}} \mathrm{QA}$ with respect to ϵ implies that the extent of any violations of the γ-orderliness, if $0 \leq \gamma \leq 1$, is bounded for any distribution with a finite second moment, e.g., for a right-skewed distribution in $\mathcal{P}_{\Upsilon}^{2}$, if $0 \leq \epsilon_1 \leq \epsilon_2 \leq \epsilon_3 \leq \frac{1}{1+\gamma}$, $\mathrm{QA}_{\epsilon_2,\gamma} \geq \mathrm{QA}_{\epsilon_3,\gamma} \geq \mathrm{QA}_{\epsilon_1,\gamma}$, then $\mathrm{QA}_{\epsilon_2,\gamma}$ will not be too far away from $\mathrm{QA}_{\epsilon_1,\gamma}$, since $\sup_{P\in\mathcal{P}_{\Upsilon}^{2}} \mathrm{QA}_{\epsilon_1,\gamma} > \sup_{P\in\mathcal{P}_{\Upsilon}^{2}} \mathrm{QA}_{\epsilon_2,\gamma} > \sup_{P\in\mathcal{P}_{\Upsilon}^{2}} \mathrm{QA}_{\epsilon_3,\gamma}$. Moreover, a stricter bound can be established for unimodal distributions. The violation of νth γ-orderliness, when $\nu \geq 2$, is also bounded, since the QA function is bounded, the νth γ-orderliness corresponds to the higher-order derivatives of the QA function with respect to ϵ.

Data Availability. Data for Figure **??** are given in SI Dataset S1. All codes have been deposited in GitHub.

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

1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae.* (Henricus Dieterich), (1823).
2. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
3. P Daniell, Observations weighted according to order. *Am. J. Math.* **42**, 222–236 (1920).
4. JW Tukey, A survey of sampling from contaminated distributions in *Contributions to probability and statistics.* (Stanford University Press), pp. 448–485 (1960).
5. WJ Dixon, Simplified Estimation from Censored Normal Samples. *The Annals Math. Stat.* **31**, 385 – 391 (1960).
6. K Danielak, T Rychlik, Theory & methods: Exact bounds for the bias of trimmed means. *Aust. & New Zealand J. Stat.* **45**, 83–96 (2003).
7. M Bieniek, Comparison of the bias of trimmed and winsorized means. *Commun. Stat. Methods* **45**, 6641–6650 (2016).
8. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
9. F Wilcoxon, Individual comparisons by ranking methods. *Biom. Bull.* **1**, 80–83 (1945).
10. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
11. Q Sun, WX Zhou, J Fan, Adaptive huber regression. *J. Am. Stat. Assoc.* **115**, 254–265 (2020).
12. DL Donoho, PJ Huber, The notion of breakdown point. *A festschrift for Erich L. Lehmann* **157184** (1983).
13. CM Stein, Efficient nonparametric testing and estimation in *Proceedings of the third Berkeley symposium on mathematical statistics and probability.* Vol. 1, pp. 187–195 (1956).
14. PJ Bickel, On adaptive estimation. *The Annals Stat.* **10**, 647–671 (1982).
15. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparametric models.* (Springer) Vol. 4, (1993).
16. JT Runnenburg, Mean, median, mode. *Stat. Neerlandica* **32**, 73–79 (1978).
17. Wv Zwet, Mean, median, mode ii. *Stat. Neerlandica* **33**, 1–5 (1979).
18. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
19. 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).
20. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
21. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).
22. 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).