

# Semiparametric robust mean estimations based on the orderliness of quantile averages

Tuban Lee

This manuscript was compiled on May 26, 2023

**As one of the most fundamental problem 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. This study establishes several forms of orderliness among quantile averages, similar to the mean-median-mode inequality, within a wide range of semiparametric distributions. From this, a sequence of advanced robust mean estimators emerges, which also explains 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,  $|m - \mu| \leq \sqrt{\frac{3}{4}}\omega$  and  $\sigma \leq \omega \leq 2\sigma$ , where  $\mu$  is the population mean,  $m$  is the population median,  $\omega$  is the root mean square deviation from the mode, and  $\sigma$  is the population standard deviation. This pioneering work revealed that despite potential bias in robust mean estimates, the deviation remains bounded in units 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  $m$  has the smallest maximum distance to  $\mu$  among all symmetric quantile averages (SQA $_{\epsilon}$ ). Daniell, in 1920, (3) analyzed a class of estimators, linear combinations of order statistics, and identified that  $\epsilon$ -symmetric trimmed mean (STM $_{\epsilon}$ ) belongs to this class. Another popular choice, the  $\epsilon$ -symmetric Winsorized mean (SWM $_{\epsilon}$ ), 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 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 (16–18). 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 (17). 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) (19). Laforgue, Clemencon, and Bertail (2019) proposed the median of randomized means (MoRM $_{k,b}$ ) (18), wherein, rather than partitioning, an arbitrary number,  $b$ , of blocks are built independently from the sample, and showed that MoRM has a 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) (20), Bickel and Freedman (1981, 1984) (21, 22), and Helmers, Janssen, and

## 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, instead of examining 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 is 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.

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

Veraverbeke (1990) (23).

Here, the  $\epsilon, b$ -stratified mean is defined as

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

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.  $SM_{\epsilon, b=3}$  is equivalent to  $STM_{\epsilon}$ , when  $\epsilon > \frac{1}{6}$ . When  $\frac{b-1}{2\epsilon} \in \mathbb{N}$ ,  $b \bmod 2 = 1$ , the basic idea of the stratified mean 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 provided in the SI Text. Although the principle resembles that of the median of means,  $SM_{\epsilon, b, n}$  is different from  $MoM_{k=\frac{n}{b}, b}$  as it does not include the random shift. Additionally, the stratified mean differs from the mean of the sample obtained through stratified sampling methods, introduced by Neyman (1934) (24) or ranked set sampling (25), introduced by McIntyre in 1952, as these sampling methods aim to obtain more representative samples or improve the efficiency of sample estimates, but the sample means 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, 17, 19) 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 mean estimators can be deduced, which exhibit 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

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

where  $w_i$ s are the weights applied to the symmetric quantile averages according to the definition of the corresponding  $L$ -estimators. For example, for the  $\epsilon$ -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, two definitions for the  $\epsilon, \gamma$ -quantile average ( $QA(\epsilon, \gamma, n)$ ) are proposed. The first definition is:

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

and the second definition is:

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

where  $\hat{Q}_n(p)$  is the empirical quantile function;  $\gamma$  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 remainder 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

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

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

## Classifying distributions by the signs of derivatives

Let  $\mathcal{P}_\mathcal{R}$  denote the set of all continuous distribution over  $\mathbb{R}$ . Without loss of generality, the discussion of all the classes outlined below is restricted to the intersection with the nonparametric class of distributions,  $\mathcal{P}_\mathcal{R}$ . 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 (26). 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 (27) which systematically categorized many common models into three classes: parametric, nonparametric, and semiparametric. 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., for a

continuous distribution,  $(f'(x) > 0 \text{ for } x \leq M) \wedge (f'(x) < 0 \text{ for } x \geq M)$ , 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) (28, 29) endeavored to determine sufficient conditions for the inequality to hold, thereby implying the possibility of its violation. The class of distributions that satisfy the mean-median-mode inequality constitutes a subclass of  $\mathcal{P}_U$ . By analogy, a right-skewed distribution is called  $\gamma$ -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 necessary and sufficient condition below hints at the relation between the mean-median-mode inequality and the  $\gamma$ -orderliness.

**Theorem .1.** *Let  $P_\Upsilon$  represent an arbitrary distribution in the set  $\mathcal{P}_\Upsilon$ .  $P_\Upsilon$  is  $\gamma$ -ordered 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  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ .*

*Proof.* Without loss of generality, consider the case of right-skewed continuous distribution. From the definition of  $\gamma$ -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  $\delta$  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, given 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  $\gamma$ -ordered. 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 is extendable to unimodal-like distributions. Suppose there is a right-skewed continuous multimodal distribution following the mean- $\gamma$ -median-first mode inequality with several smaller modes on the right side, with the first mode,  $M_1$ , having the greatest probability density, and the  $\gamma$ -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, while a distribution following the mean- $\gamma$ -median-mode inequality may not be strictly  $\gamma$ -ordered, the inequality that defines  $\gamma$ -orderliness remains valid for most quantile averages. The mean- $\gamma$ -median-mode inequality can also indicate possible bounds for  $\gamma$  in practice, e.g., for any distributions, when  $\gamma \rightarrow \infty$ , the  $\gamma$ -median will be greater than the mean and the mode, when  $\gamma \rightarrow 0$ , the  $\gamma$ -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  $\gamma$ -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  $\gamma$ -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  $\epsilon \rightarrow 0$ . By further assuming convexity, the second  $\gamma$ -orderliness can be defined for a right-skewed distribution as follows,

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

Analogously, the  $\nu$ th  $\gamma$ -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  $\nu$ th  $\gamma$ -orderliness is referred as  $\nu$ th 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  $\nu$ th ordered and  $\nu$ th  $\gamma$ -ordered, respectively. When the shape parameter of the Weibull distribution,  $\alpha$ , is smaller than 3.258, it can be shown that the Weibull distribution belong 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  $\alpha \rightarrow \infty$ , the skewness of the Weibull distribution reaches 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  $\nu$ th  $\gamma$ -ordered if and only if the family of probability distributions is  $\nu$ th  $\gamma$ -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 is  $Q(p) = \lambda Q_0(p) + \mu$ , where  $\lambda$  is the scale parameter and  $\mu$  is the location parameter. According to the definition of the  $\nu$ th  $\gamma$ -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  $\nu$ th  $\gamma$ -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  $\gamma$ -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  $\gamma$ -symmetric distribution is  $\nu$ th  $\gamma$ -ordered.*

*Proof.* The equality 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 definition of QA, the QA function of the  $\gamma$ -symmetric distribution is a horizontal line, since  $\frac{\partial \text{QA}_{\epsilon, \gamma}}{\partial \epsilon} = \gamma Q'(\gamma\epsilon) - Q'(1-\epsilon) = 0$ . So, the  $\nu$ th order derivative of QA is always zero.  $\square$



**Theorem .4.** A symmetric distribution is a special case of a  $\gamma$ -symmetric distribution when  $\gamma = 1$ .

*Proof.* Without loss of generality, assuming continuity. 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,  $2m - Q(p) = Q(1 - p)$ .  $\square$

As a consequence of Theorem .3, and due to the property that the generalized Gaussian distribution is symmetric around the median, it is found to be  $\nu$ th ordered.

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

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

It is now straightforward to show that the Pareto distribution follows the  $\nu$ th  $\gamma$ -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 continuous distribution with a monotonic decreasing pdf is second  $\gamma$ -ordered.

*Proof.* A monotonic decreasing pdf implies  $f'(x) = F^{(2)}(x) \leq 0$ . 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 .1 and .5.  $\square$

Theorem .6 provides valuable insights into the relation between modality and 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. The number of modes and their magnitudes within a distribution are closely related to the possibility of orderliness being valid, although counterexamples can always be constructed for non-monotonic distributions. It can be easily established that the gamma distribution is second  $\gamma$ -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  $x \geq 0$ ,  $\lambda > 0$ ,  $\alpha > 0$ ,  $\Gamma$  is the gamma function, it is a product of two monotonic decreasing functions under constraints. For  $\alpha > 1$ , an analytical analysis becomes challenging. Numerical results show that the orderliness is valid until  $\alpha > 140$ , the second orderliness is valid until  $\alpha > 78$ , and the third orderliness is valid until  $\alpha > 55$  (SI Text). It is instructive to consider that when  $\alpha \rightarrow \infty$  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  $\alpha$ , since  $\frac{\partial \tilde{\mu}_3(\alpha)}{\partial \alpha} = \frac{-3\alpha-2}{2(\alpha(\alpha+1))^{3/2}} < 0$ . When  $\alpha = 55$ ,  $\tilde{\mu}_3(\alpha) = 1.027$ .

Theorefore, similar to the Weibull distribution, the parameters that let the distribution not 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  $\gamma$ -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  $\gamma$ -ordered. Moreover, if the quantile function of  $X$  satisfies  $Q^{(2)}(p) \leq 0$ , the convex transformed distribution is second  $\gamma$ -ordered.

*Proof.* Let  $\phi Q_A(\epsilon) = \frac{1}{2}(\phi(Q(\gamma\epsilon)) + \phi(Q(1 - \epsilon)))$ . Then,  $\frac{\partial \phi Q_A}{\partial \epsilon} = \frac{1}{2}(\gamma\phi'(Q(\gamma\epsilon))Q'(\gamma\epsilon) - \phi'(Q(1 - \epsilon))Q'(1 - \epsilon)) = \frac{1}{2}\gamma Q'(\gamma\epsilon)(\phi'(Q(\gamma\epsilon)) - \phi'(Q(1 - \epsilon))) \leq 0$ , since for a  $\gamma$ -symmetric distribution,  $Q(\frac{1}{1+\gamma}) - Q(\gamma\epsilon) = Q(1 - \epsilon) - Q(\frac{1}{1+\gamma})$ , differentiating both sides,  $-\gamma Q'(\gamma\epsilon) = -Q'(1 - \epsilon)$ , where  $Q'(p) \geq 0$ ,  $\phi^{(2)}(x) \geq 0$ . If further differentiating the equality,  $\gamma^2 Q^{(2)}(\gamma\epsilon) = -Q^{(2)}(1 - \epsilon)$ . Since  $\frac{\partial^{(2)} \phi Q_A}{\partial \epsilon^{(2)}} = \frac{1}{2}(\gamma^2 \phi^2(Q(\gamma\epsilon))(Q'(\gamma\epsilon))^2 + \phi^2(Q(1 - \epsilon))(Q'(1 - \epsilon))^2) + \frac{1}{2}(\gamma^2 \phi'(Q(\gamma\epsilon))(Q^2(\gamma\epsilon)) + \phi'(Q(1 - \epsilon))(Q^2(1 - \epsilon))) = \frac{1}{2}((\phi^{(2)}(Q(\gamma\epsilon)) + \phi^{(2)}(Q(1 - \epsilon)))(\gamma^2 Q'(\gamma\epsilon))^2) + \frac{1}{2}((\phi'(Q(\gamma\epsilon)) - \phi'(Q(1 - \epsilon)))\gamma^2 Q^{(2)}(\gamma\epsilon))$ . If  $Q^{(2)}(p) \leq 0$ ,  $\frac{\partial^{(2)} \phi Q_A}{\partial \epsilon^{(2)}} \geq 0$ .  $\square$

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 (30), but the most straightforward solution is the exponential transformation since the derivatives are invariably positive. 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  $Q^{(2)}(p) = -2\sqrt{2\pi}\sigma e^{2\text{erfc}^{-1}(2p)^2} \text{erfc}^{-1}(2p) \leq 0$ , where  $\sigma$  is the standard deviation of the Gaussian distribution and  $\text{erfc}$  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  $\phi$  is the non-decreasing convex function in van Zwet's trailblazing work *Convex transformations of random variables* (31) if adding an additional constraint that  $\phi'(x) \geq 0$ . Consider a near-symmetric distribution  $S$ , such that  $SQ_A_\epsilon$  as a function of  $\epsilon$  fluctuates from 0 to  $\frac{1}{2}$ , with  $\mu = m$ . By definition,  $S$  is not ordered. Let  $s$  be the pdf of  $S$ . Applying the transformation  $\phi(x)$  to  $S$  decreases  $s(Q_S(\epsilon))$ , and the decrease rate, due to the order, is much smaller for  $s(Q_S(1 - \epsilon))$ . As a consequence, as the second derivative of  $\phi(x)$  increases, eventually, after a point,  $s(Q_S(\epsilon))$  becomes greater than  $s(Q_S(1 - \epsilon))$  even if it was not previously. Thus, the  $SQ_A_\epsilon$  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 mean-median difference  $\mu - m$  as a measure of skewness after standardization in 1895 (32). Bowley (1926) proposed a measure of skewness based on the  $SQ_A$ -median difference  $SQ_A_\epsilon - m$  (33). Groeneveld and Meeden (1984) (34) generalized these measures of skewness based

on van Zwet's convex transformation (31) while exploring their properties. A distribution is called monotonically right-skewed if and only if  $\forall 0 \leq \epsilon_1 \leq \epsilon_2 \leq \frac{1}{2}$ ,  $\text{SQA}_{\epsilon_1} - m \geq \text{SQA}_{\epsilon_2} - m$ . Since  $m$  is a constant, the monotonic skewness is equivalent to the orderliness. For a nonordered distribution, the signs of  $\text{SQA}_{\epsilon} - m$  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 to be generally near-symmetric, however, counterexamples can be constructed. For example, 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  $\epsilon$  is close to  $\frac{1}{2}$ . 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 (35) 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  $\mu$  and standard deviation is  $\sigma$ , the bias upper bound of the quantile average is given in the following theorem.

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

$$\sup_{P \in \mathcal{P}_{\mu=0, \sigma=1}} \text{QA}(\epsilon, \gamma) = \frac{1}{2} \left( \sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}} + \sqrt{\frac{1-\epsilon}{\epsilon}} \right), \quad [3]$$

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 (35).  $\square$

In 2020, Bernard et al. (2) refined these bounds for unimodal distributions. They further derived the bias bound of the symmetric quantile average. Let  $\mathcal{P}_T^k$  denote the set of all continuous distributions whose moments, from the first to the  $k$ th, are all finite. Here, the bias upper bound of the quantile average,  $0 \leq \gamma < 5$ , for  $P \in \mathcal{P}_U \cap \mathcal{P}_T^2$  is given as

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

The proof based on the bias bounds of any quantile (2) and the  $\gamma \geq 5$  case are given in the SI Text. The succeeding theorems highlight their safeguarding role in defining estimators based on  $\nu$ th  $\gamma$ -orderliness. The proof of Theorem .9 is provided in the SI Text.

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

**Theorem .10.**  $\sup_{P \in \mathcal{P}_U \cap \mathcal{P}_{\mu=0, \sigma=1}} \text{QA}(\epsilon, \gamma)$  is monotonic decreasing with respect to  $\epsilon$  over the interval  $[0, \frac{1}{1+\gamma}]$ , when  $0 \leq \gamma \leq 1$ .

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

When  $\frac{1}{6} < \epsilon \leq \frac{1}{1+\gamma}$ ,  $\frac{\partial \sup \text{QA}(\epsilon, \gamma)}{\partial \epsilon} = \sqrt{3} \left( \frac{\gamma}{\sqrt{\gamma\epsilon(4-3\gamma\epsilon)^{\frac{3}{2}}}} - \frac{1}{\sqrt{1-\epsilon(3\epsilon+1)^{\frac{3}{2}}}} \right)$ . When  $\gamma = 0$ ,  $\frac{\gamma}{\sqrt{\gamma\epsilon(4-3\gamma\epsilon)^{\frac{3}{2}}}} = \frac{\sqrt{\gamma}}{\sqrt{\epsilon(4-3\gamma\epsilon)^{\frac{3}{2}}}} = 0$ ,  $\frac{\partial \sup \text{QA}(\epsilon, \gamma)}{\partial \epsilon} < 0$ . For other cases, to determine whether  $\frac{\partial \sup \text{QA}(\epsilon, \gamma)}{\partial \epsilon} \leq 0$ , 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  $0 < \gamma \leq 1$ ,  $0 < \gamma \leq \min(1, \frac{1}{\epsilon} - 1)$ . When  $\frac{1}{6} < \epsilon \leq \frac{1}{2}$ ,  $\frac{1}{\epsilon} - 1 > 1$ , so  $0 < \gamma \leq 1$ . When  $\frac{1}{2} < \epsilon < 1$ ,  $0 < \gamma \leq \frac{1}{\epsilon} - 1$ . 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  $\gamma$  is equal to the boundary point of the domain. When  $\frac{1}{6} < \epsilon \leq \frac{1}{2}$ ,  $0 < \gamma \leq 1$ , since  $h(0) \rightarrow \infty$ ,  $h(1) = 108\epsilon^3 + 64\epsilon$ , 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  $g(\epsilon) \geq 0$ , the simplified inequality is satisfied. When  $\frac{1}{2} < \epsilon < 1$ ,  $0 < \gamma \leq \frac{1}{\epsilon} - 1$ . Since

433  $h(\frac{1}{\epsilon} - 1) = 108(\frac{1}{\epsilon} - 1)\epsilon^3 + \frac{64\epsilon}{\frac{1}{\epsilon} - 1}, 108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} - 162\epsilon^2 - 8\epsilon - 1 >$   
434  $108(\frac{1}{\epsilon} - 1)\epsilon^3 + \frac{64\epsilon}{\frac{1}{\epsilon} - 1} - 162\epsilon^2 - 8\epsilon - 1 = \frac{-108\epsilon^4 + 54\epsilon^3 - 18\epsilon^2 + 7\epsilon + 1}{\epsilon - 1}.$   
435 Let  $nu(\epsilon) = -108\epsilon^4 + 54\epsilon^3 - 18\epsilon^2 + 7\epsilon + 1$ , then  $nu'(\epsilon) =$   
436  $-432\epsilon^3 + 162\epsilon^2 - 36\epsilon + 7, nu''(\epsilon) = -1296\epsilon^2 + 324\epsilon - 36 < 0.$   
437 Since  $nu'(\epsilon = \frac{1}{2}) = -\frac{49}{2} < 0, nu'(\epsilon) < 0$ . Also,  $nu(\epsilon = \frac{1}{2}) = 0$ ,  
438 so  $nu(\epsilon) \leq 0$ , the simplified inequality is also satisfied. As a  
439 result, the simplified inequality is also valid within the range  
440 of  $\frac{1}{6} < \epsilon \leq \frac{1}{1+\gamma}$ , provided that  $0 < \gamma \leq 1$ . Then, it validates  
441  $\frac{\partial \sup QA(\epsilon, \gamma)}{\partial \epsilon} \leq 0$  for the same range of  $\epsilon$  and  $\gamma$ .  
442 The first and second formulae, when  $\epsilon = \frac{1}{6}$ , are all equal  
443 to  $\frac{1}{2} \left( \sqrt{\frac{\gamma}{4 - \frac{\gamma}{2}}} + \sqrt{\frac{5}{3}} \right)$ . It follows that  $\sup QA(\epsilon, \gamma)$  is con-  
444 tinuous over  $[0, \frac{1}{1+\gamma}]$ . Hence,  $\frac{\partial \sup QA(\epsilon, \gamma)}{\partial \epsilon} \leq 0$  holds for the  
445 entire range  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ , when  $0 \leq \gamma \leq 1$ , which leads to the  
446 assertion of this theorem.  $\square$

447 For a right-skewed distribution, considering the upper  
448 bound is enough. The monotonicity of  $\sup_{P \in \mathcal{P}_T^2} QA$  with  
449 respect to  $\epsilon$  implies that the extent of any violations of the  
450  $\gamma$ -orderliness, if  $0 \leq \gamma \leq 1$ , is bounded for any distribution  
451 with a finite second moment, e.g., for a right-skewed distribu-  
452 tion in  $\mathcal{P}_T^2$ , if  $\exists 0 \leq \epsilon_1 \leq \epsilon_2 \leq \epsilon_3 \leq \frac{1}{1+\gamma}$ ,  $QA_{\epsilon_2, \gamma} \geq QA_{\epsilon_3, \gamma} \geq$   
453  $QA_{\epsilon_1, \gamma}$ ,  $QA_{\epsilon_2, \gamma}$  will not be too far away from  $QA_{\epsilon_1, \gamma}$ , since  
454  $\sup_{P \in \mathcal{P}_T^2} QA_{\epsilon_1, \gamma} > \sup_{P \in \mathcal{P}_T^2} QA_{\epsilon_2, \gamma} > \sup_{P \in \mathcal{P}_T^2} QA_{\epsilon_3, \gamma}$ , and  
455 a similar bound can be established for unimodal distributions.  
456 The  $\nu$ th  $\gamma$ -orderliness, when  $\nu \geq 2$ , corresponds to the higher  
457 order derivatives of the QA function with respect to  $\epsilon$ , so its  
458 violation is also bounded.

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

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

- 464 1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae*. (Henricus  
465 Dieterich), (1823).
- 466 2. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under  
467 partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
- 468 3. P Daniell, Observations weighted according to order. *Am. J. Math.* **42**, 222–236 (1920).
- 469 4. JW Tukey, A survey of sampling from contaminated distributions in *Contributions to probability*  
470 *and statistics*. (Stanford University Press), pp. 448–485 (1960).
- 471 5. WJ Dixon, Simplified Estimation from Censored Normal Samples. *The Annals Math. Stat.* **31**,  
472 385–391 (1960).
- 473 6. K Danielak, T Rychlik, Theory & methods: Exact bounds for the bias of trimmed means. *Aust.*  
474 *& New Zealand J. Stat.* **45**, 83–96 (2003).
- 475 7. M Bieniek, Comparison of the bias of trimmed and winsorized means. *Commun. Stat. Methods*  
476 **45**, 6641–6650 (2016).
- 477 8. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.*  
478 **34**, 598–611 (1963).
- 479 9. F Wilcoxon, Individual comparisons by ranking methods. *Biom. Bull.* **1**, 80–83 (1945).
- 480 10. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
- 481 11. Q Sun, WX Zhou, J Fan, Adaptive huber regression. *J. Am. Stat. Assoc.* **115**, 254–265 (2020).
- 482 12. T Mathieu, Concentration study of m-estimators using the influence function. *Electron. J. Stat.*  
483 **16**, 3695–3750 (2022).
- 484 13. AS Nemirovskij, DB Yudin, *Problem complexity and method efficiency in optimization*. (Wiley-  
485 Interscience), (1983).
- 486 14. MR Jerrum, LG Valiant, VV Vazirani, Random generation of combinatorial structures from a  
487 uniform distribution. *Theor. computer science* **43**, 169–188 (1986).
- 488 15. N Alon, Y Matias, M Szegedy, The space complexity of approximating the frequency moments  
489 in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. pp.  
490 20–29 (1996).
- 491 16. PL Bühlmann, Bagging, subbagging and bragging for improving some prediction algorithms  
492 in *Research report/Seminar für Statistik, Eidgenössische Technische Hochschule (ETH)*.  
493 (Seminar für Statistik, Eidgenössische Technische Hochschule (ETH), Zürich), Vol. 113,  
494 (2003).
- 495 17. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.*  
496 **44**, 2695–2725 (2016).

- 497 18. P Laforge, S Cléménçon, P Bertail, On medians of (randomized) pairwise means in *Interna-*  
498 *tional Conference on Machine Learning*. (PMLR), pp. 1272–1281 (2019).
- 499 19. E Gobet, M Lerasle, D Métivier, Mean estimation for Randomized Quasi Monte Carlo method.  
500 working paper or preprint (2022).
- 501 20. B Efron, Bootstrap methods: Another look at the jackknife. *The Annals Stat.* **7**, 1–26 (1979).
- 502 21. PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. *The annals statistics* **9**,  
503 1196–1217 (1981).
- 504 22. PJ Bickel, DA Freedman, Asymptotic normality and the bootstrap in stratified sampling. *The*  
505 *annals statistics* **12**, 470–482 (1984).
- 506 23. R Helmers, P Janssen, N Veraverbeke, *Bootstrapping U-quantiles*. (CWI. Department of  
507 Operations Research, Statistics, and System Theory [BS]), (1990).
- 508 24. J Neyman, On the two different aspects of the representative method: The method of stratified  
509 sampling and the method of purposive selection. *J. Royal Stat. Soc.* **97**, 558–606 (1934).
- 510 25. G McIntyre, A method for unbiased selective sampling, using ranked sets. *Aust. journal*  
511 *agricultural research* **3**, 385–390 (1952).
- 512 26. CM Stein, Efficient nonparametric testing and estimation in *Proceedings of the third Berkeley*  
513 *symposium on mathematical statistics and probability*. Vol. 1, pp. 187–195 (1956).
- 514 27. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparamet-*  
515 *ric models*. (Springer) Vol. 4, (1993).
- 516 28. JT Runnenburg, Mean, median, mode. *Stat. Neerlandica* **32**, 73–79 (1978).
- 517 29. Wv Zwet, Mean, median, mode ii. *Stat. Neerlandica* **33**, 1–5 (1979).
- 518 30. H Rietz, On certain properties of frequency distributions of the powers and roots of the variates  
519 of a given distribution. *Proc. Natl. Acad. Sci.* **13**, 817–820 (1927).
- 520 31. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen*. (1964).
- 521 32. K Pearson, X. contributions to the mathematical theory of evolution.—ii. skew variation in  
522 homogeneous material. *Philos. Transactions Royal Soc. London (A.)* **186**, 343–414 (1895).
- 523 33. AL Bowley, *Elements of statistics*. (King) No. 8, (1926).
- 524 34. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D*  
525 *(The Stat.)* **33**, 391–399 (1984).
- 526 35. L Li, H Shao, R Wang, J Yang, Worst-case range value-at-risk with partial information. *SIAM J.*  
527 *on Financial Math.* **9**, 190–218 (2018).