

# Robust estimations from distribution structures: I. Mean

Tuobang Li

This manuscript was compiled on November 25, 2023

**As the most fundamental problem in statistics, robust location estimation has many prominent solutions, such as the trimmed mean, 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, but the underlying mechanisms largely remain unclear. This study exploited a semiparametric method to classify distributions by the asymptotic orderliness of quantile combinations with varying breakdown points, showing their interrelations and connections to parametric distributions. Further deductions explain why the Winsorized mean typically has smaller biases compared to the trimmed mean; two sequences of semiparametric robust mean estimators emerge, particularly highlighting the superiority of the median Hodges–Lehmann mean.**

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, the potential bias of the median with respect to the mean is bounded in units of a scale parameter under certain assumptions. In 2018, Li, Shao, Wang, and Yang (2) proved the bias bound of any quantile for arbitrary continuous distributions with finite second moments. Bernard, Kazzi, and Vanduffel (2020) (3) further refined these bounds for unimodal distributions with finite second moments and extended to the bounds of symmetric quantile averages. They showed that  $m$  has the smallest maximum distance to  $\mu$  among all symmetric quantile averages (SQA <sub>$\epsilon$</sub> ). Daniell, in 1920, (4) analyzed a class of estimators, linear combinations of order statistics, and identified that the  $\epsilon$ -symmetric trimmed mean (STM <sub>$\epsilon$</sub> ) belongs to this class. Another popular choice, the  $\epsilon$ -symmetric Winsorized mean (SWM <sub>$\epsilon$</sub> ), named after Winsor and introduced by Tukey (5) and Dixon (6) 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 (7, 8). In 1963, Hodges and Lehmann (9) proposed a class of nonparametric location estimators based on rank tests and, from the Wilcoxon signed-rank statistic (10), 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 (11) 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 (11). 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 (11). Sun, Zhou, and Fan (2020) examined the concentration bounds of the Huber  $M$ -estimator (12). In 2012, Catoni proposed an  $M$ -estimator for heavy-tailed samples with finite variance (13). The concept of the median of means (MoM <sub>$k, b = \frac{n}{k}, n$</sub> ) was first introduced by Nemirovsky and Yudin (1983) in their work on stochastic optimization (14), while later was revisited in Jerrum, Valiant, and Vazirani (1986), (15) and Alon, Matias and Szegedy (1996) (16)'s works. Given its good performance even for distributions with infinite second moments, the MoM has received increasing attention over the past decade (17–20). Devroye, Lerasle, Lugosi, and Oliveira (2016) showed that MoM <sub>$k, b = \frac{n}{k}, n$</sub>  nears the optimum of sub-Gaussian mean estimation with regards to concentration bounds when the distribution has a heavy tail (18). Laforgue, Clemencon, and Bertail (2019) proposed the median of randomized means (MoRM <sub>$k, b, n$</sub> ) (19), wherein, rather than partitioning, an arbitrary number,  $b$ , of blocks are built independently from the sample, and showed that MoRM <sub>$k, b, n$</sub>  has a better non-asymptotic sub-Gaussian property compared to MoM <sub>$k, b = \frac{n}{k}, n$</sub> . In fact, asymptotically, the Hodges-Lehmann (H-L) estimator is equivalent to MoM <sub>$k=2, b = \frac{n}{k}$</sub>  and MoRM <sub>$k=2, b$</sub> , and they can be seen as the pairwise mean distribution is approximated by the sampling without replacement and bootstrap, respectively. When  $k \ll n$ , the difference between sampling with replacement and without replacement is negligible. For the asymptotic validity, readers are referred to the foundational works of Efron (1979) (21), Bickel and Freedman (1981, 1984) (22, 23), and Helmers, Janssen, and Veraverbeke (1990) (24).

## 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 combinations within a distribution is investigated. By classifying distributions through the signs of derivatives, two 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.

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

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

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

where  $X_1 \leq \dots \leq X_n$  denote the order statistics of a sample of  $n$  independent and identically distributed random variables  $X_1, \dots, X_n$ .  $b \in \mathbb{N}$ ,  $b \geq 3$ , and  $b \bmod 2 = 1$ . 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 equivalent to  $\text{STM}_{\epsilon}$ , when  $\epsilon > \frac{1}{6}$ . When  $\frac{b-1}{2\epsilon} \in \mathbb{N}$ , 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 smaller sample size and the number of sampling times are provided in the SI Text. Although the principle resembles that of the median of means,  $\text{SM}_{\epsilon, b, n}$  is different from  $\text{MoM}_{k=\frac{n}{b}, b, n}$  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) (25) or ranked set sampling (26), 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 (7, 8, 18) 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, two sets 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

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

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,

$w_i = \begin{cases} 0, & i < n\epsilon \\ 1, & i \geq n\epsilon \end{cases}$ , 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  $\epsilon, \gamma$ -quantile average ( $\text{QA}_{\epsilon, \gamma, n}$ ) 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 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

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

For any weighted average, if  $\gamma$  is omitted, it is assumed to be 1. The  $\epsilon, \gamma$ -trimmed mean ( $\text{TM}_{\epsilon, \gamma, n}$ ) 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 (27). The finite-sample upper breakdown point is the finite sample breakdown point defined by Donoho and Huber (1983) (28) and also detailed in (27). The (finite-sample) lower breakdown point is replacing the infinity symbol in these definitions with negative infinity.

## 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 (29) and proposed a necessary condition for this type of problem, a contribution later explicitly recognized as initiating the field of semiparametric statistics (30). In 1982, Bickel simplified Stein's general heuristic necessary condition (29), derived sufficient conditions, and used them in formulating adaptive estimates (30). 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 (31), 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) (32, 33) 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  $\gamma$ -orderliness for a right-skewed distribution is defined as

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

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

**Theorem .1.** *A distribution is  $\gamma$ -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  $\gamma$ -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, since the derivative of cdf is pdf.  $\square$

According to Theorem .1, if a probability distribution is right-skewed and monotonic decreasing, it will always be  $\gamma$ -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- $\gamma$ -median-first mode inequality, 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, even though a distribution following the mean- $\gamma$ -median-mode inequality may not be strictly  $\gamma$ -ordered, the inequality defining the  $\gamma$ -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, a reasonable  $\gamma$  should maintain the validity of the mean- $\gamma$ -median-mode inequality.

The definition above of  $\gamma$ -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:

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

The left-skewed case can be obtained by reversing the inequality  $\frac{\partial \text{QA}}{\partial \epsilon} \leq 0$  to  $\frac{\partial \text{QA}}{\partial \epsilon} \geq 0$  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  $\gamma$ -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  $\epsilon$  as  $\epsilon$  approaches 0. By further assuming convexity, the second  $\gamma$ -orderliness can be defined for a right-skewed distribution as follows,

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

Analogously, the  $\nu$ th  $\gamma$ -orderliness of a right-skewed distribution can be defined as  $(-1)^\nu \frac{\partial^\nu \text{QA}}{\partial \epsilon^\nu} \geq 0 \wedge \dots \wedge -\frac{\partial \text{QA}}{\partial \epsilon} \geq 0$ . If  $\gamma = 1$ , the  $\nu$ th  $\gamma$ -orderliness is referred as to  $\nu$ 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  $\frac{1}{1-\ln(2)}$ , it can be shown that the Weibull distribution belongs to  $\mathcal{P}_U \cap \mathcal{P}_O \cap \mathcal{P}_{O_2}$  (SI Text). At  $\alpha \approx 3.602$ , the Weibull distribution is symmetric, and as  $\alpha \rightarrow \infty$ , 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 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 becomes  $Q(p) = \lambda Q_0(p) + \mu$ , 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,  $Q(\gamma\epsilon) = 2Q(\frac{\gamma}{1+\gamma}) - Q(1-\epsilon)$ , 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  $\gamma$ -symmetric distribution is a horizontal line, since  $\frac{\partial QA}{\partial \epsilon} = \gamma Q'(\gamma\epsilon) - Q'(1-\epsilon) = 0$ . So, the  $\nu$ th order derivative of QA is always zero.  $\square$

**Theorem .4.** A symmetric distribution is a special case of the  $\gamma$ -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  $\nu$ th ordered, as a consequence of Theorem .3. Also, the integral of all quantile averages is not equal to the mean, unless  $\gamma = 1$ , as the left and right parts have different weights. The symmetric distribution has a unique role in that its all quantile averages are equal to the mean for a distribution with a finite mean.

**Theorem .5.** Any 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  $\nu$ th  $\gamma$ -ordered, provided that  $0 \leq \gamma \leq 1$ .

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

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

**Theorem .6.** A right-skewed distribution with a monotonic decreasing pdf is second  $\gamma$ -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  $\gamma$ -orderliness for all  $\gamma \geq 0$ , which means  $\forall 0 \leq \epsilon \leq \frac{1}{1+\gamma}$ ,  $\frac{\partial QA}{\partial \epsilon} \leq 0$ . The desired result is then derived from the proof of Theorem .5, since  $(-1)^2 \frac{\partial^2 QA}{\partial \epsilon^2} \geq 0$  for all  $\gamma \geq 0$ .  $\square$

Theorem .6 provides valuable insights into the relation between modality and second  $\gamma$ -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  $\gamma$ -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  $\gamma$ -median is always greater than the corresponding probability density of the right-hand side of the  $\gamma$ -median. So although counterexamples can always be constructed for non-monotonic distributions, the general shape of a  $\gamma$ -ordered distribution should have a single dominant mode. 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  $x \geq 0$ ,  $\lambda > 0$ ,  $\alpha > 0$ , and  $\Gamma$  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 can verify that orderliness is valid if  $\alpha < 140$ , the second orderliness is valid if  $\alpha > 81$ , and the third orderliness is valid if  $\alpha < 59$  (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  $\frac{\partial \mu_3(\alpha)}{\partial \alpha} = \frac{-3\alpha-2}{2(\alpha(\alpha+1))^{3/2}} < 0$ . When  $\alpha = 59$ ,  $\mu_3(\alpha) = 1.025$ . Therefore, 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  $\gamma$ -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  $\phi QA(\epsilon, \gamma) = \frac{1}{2}(\phi(Q(\gamma\epsilon)) + \phi(Q(1-\epsilon)))$ . Then, for all  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ ,  $\frac{\partial \phi QA}{\partial \epsilon} = \frac{1}{2}(\gamma \phi'(Q(\gamma\epsilon)) Q'(\gamma\epsilon) - \phi'(Q(1-\epsilon)) Q'(1-\epsilon)) = \frac{1}{2} \gamma Q'(\gamma\epsilon) (\phi'(Q(\gamma\epsilon)) - \phi'(Q(1-\epsilon))) \leq 0$ , since for a  $\gamma$ -symmetric distribution,  $Q(\frac{\gamma}{1+\gamma}) - Q(\gamma\epsilon) = Q(1-\epsilon) - Q(\frac{1}{1+\gamma})$ , 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  $\frac{\partial^2 \phi QA}{\partial \epsilon^2} = \frac{1}{2}(\gamma^2 \phi^2(Q(\gamma\epsilon)) (Q'(\gamma\epsilon))^2 + \phi^2(Q(1-\epsilon)) (Q'(1-\epsilon))^2) + \frac{1}{2}(\gamma^2 \phi'(Q(\gamma\epsilon)) (Q^2(\gamma\epsilon)) + \phi'(Q(1-\epsilon)) (Q^2(1-\epsilon))) =$

396  $\frac{1}{2} \left( (\phi^{(2)}(Q(\gamma\epsilon)) + \phi^{(2)}(Q(1-\epsilon))) (\gamma^2 Q'(\gamma\epsilon))^2 \right) +$   
 397  $\frac{1}{2} \left( (\phi'(Q(\gamma\epsilon)) - \phi'(Q(1-\epsilon))) \gamma^2 Q^{(2)}(\gamma\epsilon) \right)$ . If  $Q^{(2)}(p) \leq 0$ ,  
 398 for all  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ ,  $\frac{\partial^{(2)} \phi QA}{\partial \epsilon^{(2)}} \geq 0$ .  $\square$

399 An application of Theorem 7 is that the lognormal  
 400 distribution is ordered as it is exponentially transformed  
 401 from the Gaussian distribution. The quantile function of  
 402 the Gaussian distribution meets the condition  $Q_{Gau}^{(2)}(p) =$   
 403  $-2\sqrt{2}\pi\sigma e^{2\text{erfc}^{-1}(2p)^2} \text{erfc}^{-1}(2p) \leq 0$ , where  $\sigma$  is the standard  
 404 deviation of the Gaussian distribution and  $\text{erfc}$  denotes the  
 405 complementary error function. Thus, the lognormal distribu-  
 406 tion is second ordered. Numerical results suggest that it is  
 407 also third ordered, although analytically proving this result is  
 408 challenging.

409 Theorem 7 also reveals a relation between convex transfor-  
 410 mation and orderliness, since  $\phi$  is the non-decreasing convex  
 411 function in van Zwet's trailblazing work *Convex transforma-*  
 412 *tions of random variables* (34) if adding an additional con-  
 413 straint that  $\phi'(x) \geq 0$ . Consider a near-symmetric distribution  
 414  $S$ , such that the SQA( $\epsilon$ ) as a function of  $\epsilon$  fluctuates from 0  
 415 to  $\frac{1}{2}$ . By definition,  $S$  is not ordered. Let  $s$  be the pdf of  $S$ .  
 416 Applying the transformation  $\phi(x)$  to  $S$  decreases  $s(Q_S(\epsilon))$ ,  
 417 and the decrease rate, due to the order, is much smaller for  
 418  $s(Q_S(1-\epsilon))$ . As a consequence, as  $\phi^{(2)}(x)$  increases, even-  
 419 tually, after a point, for all  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ ,  $s(Q_S(\epsilon))$  becomes  
 420 greater than  $s(Q_S(1-\epsilon))$  even if it was not previously. Thus,  
 421 the SQA( $\epsilon$ ) function becomes monotonically decreasing, and  $S$   
 422 becomes ordered. Accordingly, in a family of distributions that  
 423 differ by a skewness-increasing transformation in van Zwet's  
 424 sense, violations of orderliness typically occur only when the  
 425 distribution is near-symmetric.

426 Pearson proposed using the 3 times standardized mean-  
 427 median difference,  $\frac{3(\mu-m)}{\sigma}$ , as a measure of skewness in 1895  
 428 (35). Bowley (1926) proposed a measure of skewness based on  
 429 the SQA $_{\epsilon=\frac{1}{4}}$ -median difference SQA $_{\epsilon=\frac{1}{4}} - m$  (36). Groeneveld  
 430 and Meeden (1984) (37) generalized these measures of skewness  
 431 based on van Zwet's convex transformation (34) while explor-  
 432 ing their properties. A distribution is called monotonically  
 433 right-skewed if and only if  $\forall 0 \leq \epsilon_1 \leq \epsilon_2 \leq \frac{1}{2}$ , SQA $_{\epsilon_1} - m \geq$   
 434 SQA $_{\epsilon_2} - m$ . Since  $m$  is a constant, the monotonic skewness  
 435 is equivalent to the orderliness. For a nonordered distribu-  
 436 tion, the signs of SQA $_{\epsilon} - m$  with different breakdown points  
 437 might be different, implying that some skewness measures  
 438 indicate left-skewed distribution, while others suggest right-  
 439 skewed distribution. Although it seems reasonable that such a  
 440 distribution is likely to be generally near-symmetric, counterex-  
 441 amples can be constructed. For example, first consider the  
 442 Weibull distribution, when  $\alpha > \frac{1}{1-\ln(2)}$ , it is near-symmetric  
 443 and nonordered, the non-monotonicity of the SQA function  
 444 arises when  $\epsilon$  is close to  $\frac{1}{2}$ , but if then replacing the third quar-  
 445 tile with one from a right-skewed heavy-tailed distribution  
 446 leads to a right-skewed, heavy-tailed, and nonordered distri-  
 447 bution. Therefore, the validity of robust measures of skewness  
 448 based on the SQA-median difference is closely related to the  
 449 orderliness of the distribution.

450 Remarkably, in 2018, Li, Shao, Wang, Yang (2) proved the  
 451 bias bound of any quantile for arbitrary continuous distribu-  
 452 tions with finite second moments. Here, let  $\mathcal{P}_{\mu,\sigma}$  denotes the  
 453 set of continuous distributions whose mean is  $\mu$  and standard  
 454 deviation is  $\sigma$ . 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 (2).  $\square$

In 2020, Bernard et al. (3) 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

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

The proof based on the bias bounds of any quantile (3) 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  $\nu$ th  $\gamma$ -orderliness.

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

*Proof.*  $\frac{\partial \sup QA(\epsilon, \gamma)}{\partial \epsilon} = \frac{1}{4} \left( \frac{\gamma}{\sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}(\gamma\epsilon-1)^2}} - \frac{1}{\sqrt{\frac{1}{\epsilon}-1\epsilon^2}} \right)$ . When  
 $\gamma = 0$ ,  $\frac{\partial \sup QA(\epsilon, \gamma)}{\partial \epsilon} = \frac{1}{4} \left( \frac{\sqrt{\gamma}}{\sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}(\gamma\epsilon-1)^2}} - \frac{1}{\sqrt{\frac{1}{\epsilon}-1\epsilon^2}} \right) =$   
 $-\frac{1}{\sqrt{\frac{1}{\epsilon}-1\epsilon^2}} \leq 0$ . When  $\epsilon \rightarrow 0^+$ ,

$$\lim_{\epsilon \rightarrow 0^+} \left( \frac{1}{4} \left( \frac{\gamma}{\sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}(\gamma\epsilon-1)^2}} - \frac{1}{\sqrt{\frac{1}{\epsilon}-1\epsilon^2}} \right) \right) =$$

$$\lim_{\epsilon \rightarrow 0^+} \left( \frac{1}{4} \left( \frac{\sqrt{\gamma}}{\sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}(\gamma\epsilon-1)^2}} - \frac{1}{\sqrt{\frac{1}{\epsilon}-1\epsilon^2}} \right) \right) \rightarrow -\infty. \text{ Assuming } \epsilon > 0,$$

when  $0 < \gamma \leq 1$ , to prove  $\frac{\partial \sup QA(\epsilon, \gamma)}{\partial \epsilon} \leq 0$ , it is

equivalent to showing  $\frac{\sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}(\gamma\epsilon-1)^2}}{\gamma} \geq \sqrt{\frac{1}{\epsilon}-1\epsilon^2}$ . De-

fine  $L(\epsilon, \gamma) = \frac{\sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}(\gamma\epsilon-1)^2}}{\gamma}$ ,  $R(\epsilon, \gamma) = \sqrt{\frac{1}{\epsilon}-1\epsilon^2}$ .

$$\frac{L(\epsilon, \gamma)}{\epsilon^2} = \frac{\sqrt{\frac{\gamma\epsilon}{1-\gamma\epsilon}(\gamma\epsilon-1)^2}}{\gamma\epsilon^2} = \frac{1}{\gamma} \sqrt{\frac{1}{\frac{1}{\epsilon}-1}} \left( \gamma - \frac{1}{\epsilon} \right)^2, \quad \frac{R(\epsilon, \gamma)}{\epsilon^2} =$$

$$\sqrt{\frac{1}{\epsilon}-1}. \text{ Then, } \frac{L(\epsilon, \gamma)}{\epsilon^2} \geq \frac{R(\epsilon, \gamma)}{\epsilon^2} \Leftrightarrow \frac{1}{\gamma} \sqrt{\frac{1}{\frac{1}{\epsilon}-1}} \left( \gamma - \frac{1}{\epsilon} \right)^2 \geq$$

$$\sqrt{\frac{1}{\epsilon}-1} \Leftrightarrow \frac{1}{\gamma} \left( \gamma - \frac{1}{\epsilon} \right)^2 \geq \sqrt{\frac{1}{\epsilon}-1} \sqrt{\frac{1}{\epsilon}-1}. \text{ Let}$$

$$LmR\left(\frac{1}{\epsilon}\right) = \frac{1}{\gamma^2} \left( \gamma - \frac{1}{\epsilon} \right)^4 - \left( \frac{1}{\epsilon} - 1 \right) \left( \frac{1}{\epsilon} - 1 \right). \quad \frac{\partial LmR(1/\epsilon)}{\partial (1/\epsilon)} =$$

$$-\frac{4(\gamma-\frac{1}{\epsilon})^3}{\gamma^2} - \frac{1}{\epsilon} - 1 - \frac{1}{\gamma} + 1 = \frac{-4\gamma^3 + \gamma^2 + \gamma + 4\frac{1}{\epsilon^3} - 12\gamma\frac{1}{\epsilon^2} + 12\gamma^2\frac{1}{\epsilon} - 2\gamma\frac{1}{\epsilon}}{\gamma^2}.$$

Since  $0 \leq \gamma \leq 1$ ,  $0 \leq \epsilon \leq \frac{1}{1+\gamma} \Leftrightarrow 0 \leq \gamma \leq \frac{1}{\epsilon} - 1 \Leftrightarrow 1 - \frac{1}{\epsilon} \leq$

$-\gamma \leq 0 \Leftrightarrow 1 \leq \frac{1}{\epsilon} - \gamma \leq \frac{1}{\epsilon}$ . The numerator of  $\frac{\partial LmR(1/\epsilon)}{\partial (1/\epsilon)}$  can

be simplified as  $-4\gamma^3 + \gamma^2 + \gamma + 4\frac{1}{\epsilon^3} - 12\gamma\frac{1}{\epsilon^2} + 12\gamma^2\frac{1}{\epsilon} - 2\gamma\frac{1}{\epsilon} =$

$4\left(\frac{1}{\epsilon} - \gamma\right)^3 + \gamma^2 + \gamma - 2\gamma\frac{1}{\epsilon} = 4\left(\frac{1}{\epsilon} - \gamma\right)^3 - \gamma^2 + \gamma - 2\gamma\left(\frac{1}{\epsilon} - \gamma\right) =$

$\gamma(1-\gamma) + 2\left(\frac{1}{\epsilon} - \gamma\right)\left(2\left(\frac{1}{\epsilon} - \gamma\right)^2 - \gamma\right)$ . Since  $2\left(\frac{1}{\epsilon} - \gamma\right)^2 \geq 2$ ,

2  $(\frac{1}{\epsilon} - \gamma)^2 - \gamma \geq 2$ . Also,  $\gamma(1 - \gamma) \geq 0$ ,  $(\frac{1}{\epsilon} - \gamma) \geq 0$ ,  
therefore,  $\gamma(1 - \gamma) + 2(\frac{1}{\epsilon} - \gamma)(2(\frac{1}{\epsilon} - \gamma)^2 - \gamma) \geq 0$ ,  
 $\frac{\partial LmR(1/\epsilon)}{\partial(1/\epsilon)} \geq 0$ . Also,  $LmR(1 + \gamma) = \frac{1}{\gamma^2}(\gamma - 1 - \gamma)^4 -$   
 $(1 + \gamma - 1)(\frac{1}{\gamma}(1 + \gamma) - 1) = \frac{1}{\gamma^2} \geq 0$ . Therefore,  
 $LmR(\frac{1}{\epsilon}) \geq 0$  for  $\epsilon \in (0, \frac{1}{1+\gamma}]$ , provided that  
 $0 < \gamma \leq 1$ . Consequently, the simplified inequality  
 $\frac{1}{\gamma}(\gamma - \frac{1}{\epsilon})^2 \geq \sqrt{\frac{1}{\epsilon} - 1}\sqrt{\frac{1}{\gamma} - 1}$  is valid.  $\frac{\partial \sup QA(\epsilon, \gamma)}{\partial \epsilon}$  is  
non-positive throughout the interval  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ , given that  
 $0 \leq \gamma \leq 1$ , the proof is complete.  $\square$

**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  $\epsilon$  on the interval  $[0, \frac{1}{1+\gamma}]$ , provided  
that  $0 \leq \gamma \leq 1$ .

*Proof.* When  $0 \leq \epsilon \leq \frac{1}{6}$ ,  $\frac{\partial \sup QA}{\partial \epsilon} = \frac{\gamma}{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}(4-3\epsilon\gamma)^2}} -$   
 $\frac{1}{3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}} = \frac{\sqrt{\gamma}}{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}(4-3\epsilon\gamma)^2}} - \frac{1}{3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}}$ . If  $\gamma = 0$   
and  $\epsilon \rightarrow 0^+$ ,  $\frac{\partial \sup QA}{\partial \epsilon} = -\frac{1}{3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}} < 0$ . If

$\epsilon \rightarrow 0^+$ ,  $\lim_{\epsilon \rightarrow 0^+} \left( \frac{\gamma}{(4-3\gamma\epsilon)^2 \sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}} - \frac{1}{3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}} \right) =$   
 $\lim_{\epsilon \rightarrow 0^+} \left( \frac{\sqrt{3\gamma}}{\sqrt{43}\epsilon} - \frac{1}{6\sqrt{\epsilon^3}} \right) \rightarrow -\infty$ , for all  $0 \leq \gamma \leq 1$ ,  
so,  $\frac{\partial \sup QA}{\partial \epsilon} < 0$ . When  $0 < \epsilon \leq \frac{1}{6}$  and  
 $0 < \gamma \leq 1$ , to prove  $\frac{\partial \sup QA}{\partial \epsilon} \leq 0$ , it is equivalent

to showing  $\frac{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}(4-3\epsilon\gamma)^2}}{3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}} \geq 3\sqrt{\frac{4}{\epsilon}-9\epsilon^2}$ . 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  $LmR(\frac{1}{\epsilon}) =$   
 $\frac{1}{\gamma^2} \left( \frac{4}{\epsilon} - 3\gamma \right)^4 - 9 \left( \frac{12}{\epsilon\gamma} - 9 \right) \left( \frac{4}{\epsilon} - 9 \right)$ .  $\frac{\partial LmR(1/\epsilon)}{\partial(1/\epsilon)} = \frac{16(\frac{4}{\epsilon}-3\gamma)^3}{\gamma^2} -$   
 $36 \left( \frac{12}{\epsilon\gamma} - 9 \right) - \frac{108(4\frac{4}{\epsilon}-9)}{\gamma} = \frac{4(4(\frac{4}{\epsilon}-3\gamma)^3 - 27\gamma(\frac{4}{\epsilon}-3\gamma) + 27(9-\frac{4}{\epsilon}))}{\gamma^2} =$   
 $\frac{4(256\frac{1}{\epsilon^3} - 576\frac{1}{\epsilon^2}\gamma + 432\frac{1}{\epsilon}\gamma^2 - 216\frac{1}{\epsilon}\gamma - 108\gamma^3 + 81\gamma^2 + 243\gamma)}{\gamma^2}$ . Since

$256\frac{1}{\epsilon^3} - 576\frac{1}{\epsilon^2}\gamma + 432\frac{1}{\epsilon}\gamma^2 - 216\frac{1}{\epsilon}\gamma - 108\gamma^3 + 81\gamma^2 + 243\gamma \geq$   
 $1536\frac{1}{\epsilon^2} - 576\frac{1}{\epsilon} + 432\frac{1}{\epsilon}\gamma^2 - 216\frac{1}{\epsilon}\gamma - 108\gamma^3 + 81\gamma^2 + 243\gamma \geq$   
 $924\frac{1}{\epsilon^2} + 36\frac{1}{\epsilon} - 216\frac{1}{\epsilon} + 432\frac{1}{\epsilon}\gamma^2 - 108\gamma^3 + 81\gamma^2 + 243\gamma \geq$   
 $924\frac{1}{\epsilon^2} + 36\frac{1}{\epsilon} - 216\frac{1}{\epsilon} + 513\gamma^2 - 108\gamma^3 + 243\gamma > 0$ ,  
 $\frac{\partial LmR(1/\epsilon)}{\partial(1/\epsilon)} > 0$ . Also,  $LmR(6) = \frac{81(\gamma-8)((\gamma-8)^3+15\gamma)}{\gamma^2} >$   
 $0 \Leftrightarrow \gamma^4 - 32\gamma^3 + 399\gamma^2 - 2168\gamma + 4096 > 0$ . 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,  $LmR(\frac{1}{\epsilon}) \geq 0$  for  $\epsilon \in (0, \frac{1}{6}]$ , if  $0 < \gamma \leq 1$ . The first  
part is finished.

When  $\frac{1}{6} < \epsilon \leq \frac{1}{1+\gamma}$ ,  $\frac{\partial \sup QA}{\partial \epsilon} =$   
 $\sqrt{3} \left( \frac{\gamma}{\sqrt{\gamma\epsilon(4-3\gamma\epsilon)^{\frac{3}{2}}}} - \frac{1}{\sqrt{1-\epsilon}(3\epsilon+1)^{\frac{3}{2}}} \right)$ . 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  $\frac{\partial \sup QA}{\partial \epsilon} = \sqrt{3} \left( -\frac{1}{\sqrt{1-\epsilon}(3\epsilon+1)^{\frac{3}{2}}} \right) < 0$ ,

for all  $\frac{1}{6} < \epsilon \leq \frac{1}{1+\gamma}$ . If  $\gamma > 0$ , to determine whether  
 $\frac{\partial \sup QA}{\partial \epsilon} \leq 0$ , 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  $\gamma$  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  $\gamma$  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  $g(\epsilon) \geq 0$ ,  $108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} - 162\epsilon^2 - 8\epsilon - 1 \geq 0$ .  
When  $\frac{1}{2} \leq \epsilon < 1$ ,  $0 < \gamma \leq \frac{1}{\epsilon} - 1$ . Since  
 $h(\frac{1}{\epsilon} - 1) = 108(\frac{1}{\epsilon} - 1)\epsilon^3 + \frac{64\epsilon}{\frac{1}{\epsilon} - 1}$ ,  $108\gamma\epsilon^3 + \frac{64\epsilon}{\gamma} - 162\epsilon^2 - 8\epsilon - 1 >$   
 $108(\frac{1}{\epsilon} - 1)\epsilon^3 + \frac{64\epsilon}{\frac{1}{\epsilon} - 1} - 162\epsilon^2 - 8\epsilon - 1 = \frac{-108\epsilon^4 + 54\epsilon^3 - 18\epsilon^2 + 7\epsilon + 1}{\epsilon - 1}$ .  
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  
 $\frac{\partial \sup QA}{\partial \epsilon} \leq 0$  for the same range of  $\epsilon$  and  $\gamma$ .

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  $\sup QA(\epsilon, \gamma)$  is contin-  
uous over  $[0, \frac{1}{1+\gamma}]$ . Hence,  $\frac{\partial \sup QA}{\partial \epsilon} \leq 0$  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  $\mathcal{P}_\Gamma^k$  denote the set of all continuous distributions whose  
moments, from the first to the  $k$ th, are all finite. For a  
right-skewed distribution, it suffices to consider the upper  
bound. The monotonicity of  $\sup_{P \in \mathcal{P}_\Gamma^2} QA$  with respect to  $\epsilon$   
implies that the extent of any violations of the  $\gamma$ -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  $\mathcal{P}_\Gamma^2$ , if  $0 \leq \epsilon_1 \leq \epsilon_2 \leq \epsilon_3 \leq \frac{1}{1+\gamma}$ ,  $QA_{\epsilon_2, \gamma} \geq QA_{\epsilon_3, \gamma} \geq$   
 $QA_{\epsilon_1, \gamma}$ , then  $QA_{\epsilon_2, \gamma}$  will not be too far away from  $QA_{\epsilon_1, \gamma}$ ,  
since  $\sup_{P \in \mathcal{P}_\Gamma^2} QA_{\epsilon_1, \gamma} > \sup_{P \in \mathcal{P}_\Gamma^2} QA_{\epsilon_2, \gamma} > \sup_{P \in \mathcal{P}_\Gamma^2} QA_{\epsilon_3, \gamma}$ .  
Moreover, a stricter bound can be established for unimodal  
distributions according to Bernard et al. 's result (3). The  
violation of  $\nu$ th  $\gamma$ -orderliness, when  $\nu \geq 2$ , is also bounded,  
since the QA function is bounded, the  $\nu$ th  $\gamma$ -orderliness cor-  
responds to the higher-order derivatives of the QA function  
with respect to  $\epsilon$ .



## Congruent Distribution

In the realm of nonparametric statistics, the relative differences, or orders, of robust estimators are of primary importance. A key implication of this principle is that when there is a shift in the parameters of the underlying distribution, all nonparametric estimates should asymptotically change in the same direction, if they are estimating the same attribute of the distribution. If, on the other hand, the mean suggests an increase in the location of the distribution while the median indicates a decrease, a contradiction arises. It is worth noting that such contradiction is not possible for any  $LL$ -statistics in a location-scale distribution, as explained in Theorem .2 and .20. However, it is possible to construct counterexamples to the aforementioned implication in a shape-scale distribution. In the case of the Weibull distribution, its quantile function is  $Q_{Wei}(p) = \lambda(-\ln(1-p))^{1/\alpha}$ , where  $0 \leq p \leq 1$ ,  $\alpha > 0$ ,  $\lambda > 0$ ,  $\lambda$  is a scale parameter,  $\alpha$  is a shape parameter,  $\ln$  is the natural logarithm function. Then,  $m = \lambda \sqrt[\alpha]{\ln(2)}$ ,  $\mu = \lambda \Gamma(1 + \frac{1}{\alpha})$ , where  $\Gamma$  is the gamma function. When  $\alpha = 1$ ,  $m = \lambda \ln(2) \approx 0.693\lambda$ ,  $\mu = \lambda$ , when  $\alpha = \frac{1}{2}$ ,  $m = \lambda \ln^2(2) \approx 0.480\lambda$ ,  $\mu = 2\lambda$ , the mean increases as  $\alpha$  changes from 1 to  $\frac{1}{2}$ , but the median decreases. In the last section, the fundamental role of quantile average was demonstrated by using the method of classifying distributions through the signs of derivatives. To avoid such scenarios, this method can also be used. Let the quantile average function of a parametric distribution be denoted as QA( $\epsilon, \gamma, \alpha_1, \dots, \alpha_i, \dots, \alpha_k$ ), where  $\alpha_i$  represent the parameters of the distribution, then, a distribution is  $\gamma$ -congruent if and only if the sign of  $\frac{\partial QA}{\partial \alpha_i}$  remains the same for all  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ . If  $\frac{\partial QA}{\partial \alpha_i}$  is equal to zero or undefined, it can be considered both positive and negative, and thus does not impact the analysis. A distribution is completely  $\gamma$ -congruent if and only if it is  $\gamma$ -congruent and all its central moment kernel distributions are also  $\gamma$ -congruent. Setting  $\gamma = 1$  constitutes the definitions of congruence and complete congruence. Replacing the QA with  $\gamma mHLM$  (defined in the following section) gives the definition of  $\gamma$ - $U$ -congruence. Chebyshev's inequality implies that, for any probability distributions with finite second moments, as the parameters change, even if some  $LL$ -statistics change in a direction different from that of the population mean, the magnitude of the changes in the  $LL$ -statistics remains bounded compared to the changes in the population mean. Furthermore, distributions with infinite moments can be  $\gamma$ -congruent, since the definition is based on the quantile average, not the population mean.

The following theorems show the conditions that a distribution is congruent or  $\gamma$ -congruent.

**Theorem .11.** *A symmetric distribution is always congruent and  $U$ -congruent.*

*Proof.* As shown in Theorem .3 and Theorem .19, for any symmetric distribution, all quantile averages and all  $\gamma mHLM$ s coincide. The conclusion follows immediately.  $\square$

**Theorem .12.** *A positive definite location-scale distribution is always  $\gamma$ -congruent.*

*Proof.* As shown in Theorem .2, for a location-scale distribution, any quantile average can be expressed as  $\lambda QA_0(\epsilon, \gamma) + \mu$ . Therefore, the derivatives with respect to the parameters  $\lambda$

or  $\mu$  are always positive. By application of the definition, the desired outcome is obtained.  $\square$

For the Pareto distribution,  $\frac{\partial Q}{\partial \alpha} = \frac{x_m(1-p)^{-1/\alpha} \ln(1-p)}{\alpha^2}$ . Since  $\ln(1-p) < 0$  for all  $0 < p < 1$ ,  $(1-p)^{-1/\alpha} > 0$  for all  $0 < p < 1$  and  $\alpha > 0$ , so  $\frac{\partial Q}{\partial \alpha} < 0$ , and therefore  $\frac{\partial QA}{\partial \alpha} < 0$ , the Pareto distribution is  $\gamma$ -congruent. It is also  $\gamma$ - $U$ -congruent, since  $\gamma mHLM$  can also express as a function of  $Q(p)$ . For the lognormal distribution,  $\frac{\partial QA}{\partial \sigma} = \frac{1}{2} \left( \sqrt{2} \operatorname{erfc}^{-1}(2\gamma\epsilon) \left( -e^{\frac{\sqrt{2}\mu - 2\sigma \operatorname{erfc}^{-1}(2\gamma\epsilon)}{\sqrt{2}}} \right) + \left( -\sqrt{2} \right) \operatorname{erfc}^{-1}(2(1-\epsilon)) e^{\frac{\sqrt{2}\mu - 2\sigma \operatorname{erfc}^{-1}(2(1-\epsilon))}{\sqrt{2}}} \right)$ . Since the inverse complementary error function is positive when the input is smaller than 1, and negative when the input is larger than 1, and symmetry around 1, if  $0 \leq \gamma \leq 1$ ,  $\operatorname{erfc}^{-1}(2\gamma\epsilon) \geq -\operatorname{erfc}^{-1}(2-2\epsilon)$ ,  $e^{\mu - \sqrt{2}\sigma \operatorname{erfc}^{-1}(2-2\epsilon)} > e^{\mu - \sqrt{2}\sigma \operatorname{erfc}^{-1}(2\gamma\epsilon)}$ . Therefore, if  $0 \leq \gamma \leq 1$ ,  $\frac{\partial QA}{\partial \sigma} > 0$ , the lognormal distribution is  $\gamma$ -congruent. Theorem .11 implies that the generalized Gaussian distribution is congruent and  $U$ -congruent. For the Weibull distribution, when  $\alpha$  changes from 1 to  $\frac{1}{2}$ , the average probability density on the left side of the median increases, since  $\frac{\frac{1}{2}}{\lambda \ln(2)} < \frac{\frac{1}{2}}{\lambda \ln^2(2)}$ , but the mean increases, indicating that the distribution is more heavy-tailed, the probability density of large values will also increase. So, the reason for non-congruence of the Weibull distribution lies in the simultaneous increase of probability densities on two opposite sides as the shape parameter changes: one approaching the bound zero and the other approaching infinity. Note that the gamma distribution does not have this issue, Numerical results indicate that it is likely to be congruent.

## The Impact of $\gamma$ -Orderliness on Weighted Inequalities

Analogous to the  $\gamma$ -orderliness, the  $\gamma$ -trimming inequality for a right-skewed distribution is defined as  $\forall 0 \leq \epsilon_1 \leq \epsilon_2 \leq \frac{1}{1+\gamma}$ ,  $TM_{\epsilon_1, \gamma} \geq TM_{\epsilon_2, \gamma}$ .  $\gamma$ -orderliness is a sufficient condition for the  $\gamma$ -trimming inequality, as proven in the SI Text. The next theorem shows a relation between the  $\epsilon, \gamma$ -quantile average and the  $\epsilon, \gamma$ -trimmed mean under the  $\gamma$ -trimming inequality, suggesting the  $\gamma$ -orderliness is not a necessary condition for the  $\gamma$ -trimming inequality.

**Theorem .13.** *For a distribution that is right-skewed and follows the  $\gamma$ -trimming inequality, it is asymptotically true that the quantile average is always greater or equal to the corresponding trimmed mean with the same  $\epsilon$  and  $\gamma$ , for all  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ .*

*Proof.* According to the definition of the  $\gamma$ -trimming inequality:  $\forall 0 \leq \epsilon \leq \frac{1}{1+\gamma}$ ,  $\frac{1}{1-\epsilon-\gamma\epsilon+2\delta} \int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u) du \geq \frac{1}{1-\epsilon-\gamma\epsilon} \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du$ , where  $\delta$  is an infinitesimal positive quantity. Subsequently, rewriting the inequality gives  $\int_{\gamma\epsilon-\delta}^{1-\epsilon+\delta} Q(u) du - \frac{1-\epsilon-\gamma\epsilon+2\delta}{1-\epsilon-\gamma\epsilon} \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du \geq 0 \Leftrightarrow \int_{1-\epsilon}^{1-\epsilon+\delta} Q(u) du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u) du - \frac{2\delta}{1-\epsilon-\gamma\epsilon} \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du \geq 0$ . Since  $\delta \rightarrow 0^+$ ,  $\frac{1}{2\delta} \left( \int_{1-\epsilon}^{1-\epsilon+\delta} Q(u) du + \int_{\gamma\epsilon-\delta}^{\gamma\epsilon} Q(u) du \right) = \frac{Q(\gamma\epsilon) + Q(1-\epsilon)}{2} \geq \frac{1}{1-\epsilon-\gamma\epsilon} \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du$ , the proof is complete.  $\square$

An analogous result about the relation between the  $\epsilon, \gamma$ -trimmed mean and the  $\epsilon, \gamma$ -Winsorized mean can be obtained in the following theorem.

**Theorem .14.** *For a right-skewed distribution following the  $\gamma$ -trimming inequality, asymptotically, the Winsorized mean is always greater or equal to the corresponding trimmed mean with the same  $\epsilon$  and  $\gamma$ , for all  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ , provided that  $0 \leq \gamma \leq 1$ . If assuming  $\gamma$ -orderliness, the inequality is valid for any non-negative  $\gamma$ .*

*Proof.* According to Theorem .13,  $\frac{Q(\gamma\epsilon) + Q(1-\epsilon)}{2} \geq \frac{1}{1-\epsilon-\gamma\epsilon} \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du \Leftrightarrow \gamma\epsilon(Q(\gamma\epsilon) + Q(1-\epsilon)) \geq (\frac{2\gamma\epsilon}{1-\epsilon-\gamma\epsilon}) \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du$ . Then, if  $0 \leq \gamma \leq 1, (1 - \frac{1}{1-\epsilon-\gamma\epsilon}) \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du + \gamma\epsilon(Q(\gamma\epsilon) + Q(1-\epsilon)) \geq 0 \Rightarrow \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du + \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon) \geq \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du + \gamma\epsilon(Q(\gamma\epsilon) + Q(1-\epsilon)) \geq \frac{1}{1-\epsilon-\gamma\epsilon} \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du$ , the proof of the first assertion is complete. The second assertion is established in Theorem 0.3. in the SI Text.  $\square$

Replacing the TM in the  $\gamma$ -trimming inequality with WA forms the definition of the  $\gamma$ -weighted inequality. The  $\gamma$ -orderliness also implies the  $\gamma$ -Winsorization inequality when  $0 \leq \gamma \leq 1$ , as proven in the SI Text. The same rationale as presented in Theorem .2, for a location-scale distribution characterized by a location parameter  $\mu$  and a scale parameter  $\lambda$ , asymptotically, any  $WA(\epsilon, \gamma)$  can be expressed as  $\lambda WA_0(\epsilon, \gamma) + \mu$ , where  $WA_0(\epsilon, \gamma)$  is an function of  $Q_0(p)$  according to the definition of the weighted average. Adhering to the rationale present in Theorem .2, for any probability distribution within a location-scale family, a necessary and sufficient condition for whether it follows the  $\gamma$ -weighted inequality is whether the family of probability distributions also adheres to the  $\gamma$ -weighted inequality.

To construct weighted averages based on the  $\nu$ th  $\gamma$ -orderliness and satisfying the corresponding weighted inequality, when  $0 \leq \gamma \leq 1$ , let  $\mathcal{B}_i = \int_{i\epsilon}^{(i+1)\epsilon} QA(u, \gamma) du$ ,  $ka = k\epsilon + c$ . From the  $\gamma$ -orderliness for a right-skewed distribution, it follows that,  $-\frac{\partial QA}{\partial \epsilon} \geq 0 \Leftrightarrow \forall 0 \leq a \leq 2a \leq \frac{1}{1+\gamma}, -\frac{(QA(2a, \gamma) - QA(a, \gamma))}{a} \geq 0 \Rightarrow \mathcal{B}_i - \mathcal{B}_{i+1} \geq 0$ , if  $0 \leq \gamma \leq 1$ . Suppose that  $\mathcal{B}_i = \mathcal{B}_0$ . Then, the  $\epsilon, \gamma$ -block Winsorized mean, is defined as

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

which is double weighting the leftest and rightest blocks having sizes of  $\gamma\epsilon n$  and  $\epsilon n$ , respectively. As a consequence of  $\mathcal{B}_i - \mathcal{B}_{i+1} \geq 0$ , the  $\gamma$ -block Winsorization inequality is valid, provided that  $0 \leq \gamma \leq 1$ . The block Winsorized mean uses two blocks to replace the trimmed parts, not two single quantiles. The subsequent theorem provides an explanation for this difference.

**Theorem .15.** *Asymptotically, for a right-skewed distribution following the  $\gamma$ -orderliness, the Winsorized mean is always greater than or equal to the corresponding block Winsorized mean with the same  $\epsilon$  and  $\gamma$ , for all  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ , provided that  $0 \leq \gamma \leq 1$ .*

*Proof.* From the definitions of BWM and WM, the statement necessitates  $\int_{\gamma\epsilon}^{1-\epsilon} Q(u) du + \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon) \geq \int_{\gamma\epsilon}^{1-\epsilon} Q(u) du + \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u) du + \int_{1-2\epsilon}^{1-\epsilon} Q(u) du \Leftrightarrow \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon) \geq \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u) du + \int_{1-2\epsilon}^{1-\epsilon} Q(u) du$ . Define  $WML(x) = Q(\gamma\epsilon)$  and  $BWML(x) = Q(x)$ . In both functions, the interval for  $x$  is specified as  $[\gamma\epsilon, 2\gamma\epsilon]$ . Then, define  $WMu(y) = Q(1-\epsilon)$  and  $BWMu(y) = Q(y)$ . In both functions, the interval for  $y$  is specified as  $[1-2\epsilon, 1-\epsilon]$ . The function  $y : [\gamma\epsilon, 2\gamma\epsilon] \rightarrow [1-2\epsilon, 1-\epsilon]$  defined by  $y(x) = 1 - \frac{x}{\gamma}$  is a bijection.  $WML(x) + WMu(y(x)) = Q(\gamma\epsilon) + Q(1-\epsilon) \geq BWML(x) + BWMu(y(x)) = Q(x) + Q(1 - \frac{x}{\gamma})$  is valid for all  $x \in [\gamma\epsilon, 2\gamma\epsilon]$ , according to the definition of  $\gamma$ -orderliness. Integration of the left side yields,  $\int_{\gamma\epsilon}^{2\gamma\epsilon} (WML(u) + WMu(y(u))) du = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(\gamma\epsilon) du + \int_{y(\gamma\epsilon)}^{y(2\gamma\epsilon)} Q(1-\epsilon) du = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(\gamma\epsilon) du + \int_{1-2\epsilon}^{1-\epsilon} Q(1-\epsilon) du = \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon)$ , while integration of the right side yields  $\int_{\gamma\epsilon}^{2\gamma\epsilon} (BWML(x) + BWMu(y(x))) dx = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u) du + \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(1 - \frac{x}{\gamma}) dx = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u) du + \int_{1-2\epsilon}^{1-\epsilon} Q(u) du$ , which are the left and right sides of the desired inequality. Given that the upper limits and lower limits of the integrations are different for each term, the condition  $0 \leq \gamma \leq 1$  is necessary for the desired inequality to be valid.  $\square$

From the second  $\gamma$ -orderliness for a right-skewed distribution,  $\frac{\partial^2 QA}{\partial^2 \epsilon} \geq 0 \Rightarrow \forall 0 \leq a \leq 2a \leq 3a \leq \frac{1}{1+\gamma}, \frac{1}{a} \left( \frac{(QA(3a, \gamma) - QA(2a, \gamma))}{a} - \frac{(QA(2a, \gamma) - QA(a, \gamma))}{a} \right) \geq 0 \Rightarrow$  if  $0 \leq \gamma \leq 1, \mathcal{B}_i - 2\mathcal{B}_{i+1} + \mathcal{B}_{i+2} \geq 0$ .  $SM_\epsilon$  can thus be interpreted as assuming  $\gamma = 1$  and replacing the two blocks,  $\mathcal{B}_i + \mathcal{B}_{i+2}$  with one block  $2\mathcal{B}_{i+1}$ . From the  $\nu$ th  $\gamma$ -orderliness for a right-skewed distribution, the recurrence relation of the derivatives naturally produces the alternating binomial coefficients,

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

Based on the  $\nu$ th orderliness, the  $\epsilon, \gamma$ -binomial mean is introduced as

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

where  $\mathfrak{B}_{i,j} = \sum_{l=n\gamma\epsilon(j+(i-1)(\nu+1)+1)}^{n\epsilon(j+(i-1)(\nu+1)+1)} (X_l + X_{n-l+1})$ . If  $\nu$  is not indicated, it defaults to  $\nu = 3$ . Since the alternating sum of binomial coefficients equals zero, when  $\nu \ll \epsilon^{-1}$  and  $\epsilon \rightarrow 0$ ,  $BM \rightarrow \mu$ . The solutions for the continuity of the breakdown point is the same as that in SM and not repeated here. The equalities  $BM_{\nu=1, \epsilon} = BWM_\epsilon$  and  $BM_{\nu=2, \epsilon} = SM_{\epsilon, b=3}$  hold, when  $\gamma = 1$  and their respective  $\epsilon$ s are identical. Interestingly,



the biases of the  $SM_{\epsilon=\frac{1}{3}, b=3}$  and the  $WM_{\epsilon=\frac{1}{3}}$  are nearly indistinguishable in common asymmetric unimodal distributions such as Weibull, gamma, lognormal, and Pareto (SI Dataset S1). This indicates that their robustness to departures from the symmetry assumption is practically similar under unimodality, even though they are based on different orders of orderliness. If single quantiles are used, based on the second  $\gamma$ -orderliness, the stratified quantile mean can be defined as

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

$SQM_{\epsilon=\frac{1}{4}}$  is the Tukey's midhinge (38). In fact,  $SQM$  is a subcase of  $SM$  when  $\gamma = 1$  and  $b \rightarrow \infty$ , so the solution for the continuity of the breakdown point,  $\frac{1}{\epsilon} \bmod 4 \neq 0$ , is identical. However, since the definition is based on the empirical quantile function, no decimal issues related to order statistics will arise. The next theorem explains another advantage.

**Theorem .16.** *For a right-skewed second  $\gamma$ -ordered distribution, asymptotically,  $SQM_{\epsilon, \gamma}$  is always greater or equal to the corresponding  $BM_{\nu=2, \epsilon, \gamma}$  with the same  $\epsilon$  and  $\gamma$ , for all  $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ , if  $0 \leq \gamma \leq 1$ .*

*Proof.* For simplicity, suppose the order statistics of the sample are distributed into  $\epsilon^{-1} \in \mathbb{N}$  blocks in the computation of both  $SQM_{\epsilon, \gamma}$  and  $BM_{\nu=2, \epsilon, \gamma}$ . The computation of  $BM_{\nu=2, \epsilon, \gamma}$  alternates between weighting and non-weighting, let '0' denote the block assigned with a weight of zero and '1' denote the block assigned with a weight of one, the sequence indicating the weighted or non-weighted status of each block is: 0, 1, 0, 0, 1, 0, ... Let this sequence be denoted by  $a_{BM_{\nu=2, \epsilon, \gamma}}(j)$ , its formula is  $a_{BM_{\nu=2, \epsilon, \gamma}}(j) = \lfloor \frac{j \bmod 3}{2} \rfloor$ . Similarly, the computation of  $SQM_{\epsilon, \gamma}$  can be seen as positioning quantiles ( $p$ ) at the beginning of the blocks if  $0 < p < \frac{1}{1+\gamma}$ , and at the end of the blocks if  $p > \frac{1}{1+\gamma}$ . The sequence of denoting whether each block's quantile is weighted or not weighted is: 0, 1, 0, 1, 0, 1, ... Let the sequence be denoted by  $a_{SQM_{\epsilon, \gamma}}(j)$ , the formula of the sequence is  $a_{SQM_{\epsilon, \gamma}}(j) = j \bmod 2$ . If pairing all blocks in  $BM_{\nu=2, \epsilon, \gamma}$  and all quantiles in  $SQM_{\epsilon, \gamma}$ , there are two possible pairings of  $a_{BM_{\nu=2, \epsilon, \gamma}}(j)$  and  $a_{SQM_{\epsilon, \gamma}}(j)$ . One pairing occurs when  $a_{BM_{\nu=2, \epsilon, \gamma}}(j) = a_{SQM_{\epsilon, \gamma}}(j) = 1$ , while the other involves the sequence 0, 1, 0 from  $a_{BM_{\nu=2, \epsilon, \gamma}}(j)$  paired with 1, 0, 1 from  $a_{SQM_{\epsilon, \gamma}}(j)$ . By leveraging the same principle as Theorem .15 and the second  $\gamma$ -orderliness (replacing the two quantile averages with one quantile average between them), the desired result follows.  $\square$

The biases of  $SQM_{\epsilon=\frac{1}{3}}$ , which is based on the second orderliness with a quantile approach, are notably similar to those of  $BM_{\nu=3, \epsilon=\frac{1}{8}}$ , which is based on the third orderliness with a block approach, in common asymmetric unimodal distributions (Figure 1).

## Hodges–Lehmann Inequality and $\gamma$ -U-Orderliness

The Hodges–Lehmann estimator stands out as a unique robust location estimator due to its definition being substantially dissimilar from conventional  $L$ -estimators,  $R$ -estimators, and  $M$ -estimators. In their landmark paper, *Estimates of location based on rank tests*, Hodges and Lehmann (9) proposed two methods for computing the H-L estimator: the Wilcoxon score

$R$ -estimator and the median of pairwise means. The Wilcoxon score  $R$ -estimator is a location estimator based on signed-rank test, or  $R$ -estimator, (9) and was later independently discovered by Sen (1963) (39). However, the median of pairwise means is a generalized  $L$ -statistic and a trimmed  $U$ -statistic, as classified by Serfling in his novel conceptualized study in 1984 (40). Serfling further advanced the understanding by generalizing the H-L kernel as  $hl_k(x_1, \dots, x_k) = \frac{1}{k} \sum_{i=1}^k x_i$ , where  $k \in \mathbb{N}$  (40). Here, the weighted H-L kernel is defined as  $whl_k(x_1, \dots, x_k) = \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i}$ , where  $w_i$ s are the weights applied to each element.

By using the weighted H-L kernel and the  $L$ -estimator, it is now clear that the Hodges–Lehmann estimator is an  $LL$ -statistic, the definition of which is provided as follows:

$$LL_{k, \epsilon, \gamma, n} := L_{\epsilon_0, \gamma, n} \left( \text{sort} \left( (whl_k(X_{N_1}, \dots, X_{N_k}))_{N=1}^{(n)} \right) \right),$$

where  $L_{\epsilon_0, \gamma, n}(Y)$  represents the  $\epsilon_0, \gamma$ - $L$ -estimator that uses the sorted sequence,  $\text{sort} \left( (whl_k(X_{N_1}, \dots, X_{N_k}))_{N=1}^{(n)} \right)$ , as input. The upper asymptotic breakdown point of  $LL_{k, \epsilon, \gamma}$  is  $\epsilon = 1 - (1 - \epsilon_0)^{\frac{1}{k}}$ , as proven in REDS II. There are two ways to adjust the breakdown point: either by setting  $k$  as a constant and adjusting  $\epsilon_0$ , or by setting  $\epsilon_0$  as a constant and adjusting  $k$ . In the above definition,  $k$  is discrete, but the bootstrap method can be applied to ensure the continuity of  $k$ , also making the breakdown point continuous. Specifically, if  $k \in \mathbb{R}$ , let the bootstrap size be denoted by  $b$ , then first sampling the original sample  $(1 - k + \lfloor k \rfloor)b$  times with each sample size of  $\lfloor k \rfloor$ , and then subsequently sampling  $(1 - \lfloor k \rfloor + k)b$  times with each sample size of  $\lfloor k \rfloor$ ,  $(1 - k + \lfloor k \rfloor)b \in \mathbb{N}$ ,  $(1 - \lfloor k \rfloor + k)b \in \mathbb{N}$ . The corresponding kernels are computed separately, and the pooled sorted sequence is used as the input for the  $L$ -estimator. Let  $\mathbf{S}_k$  represent the sorted sequence. Indeed, for any finite sample,  $X$ , when  $k = n$ ,  $\mathbf{S}_k$  becomes a single point,  $whl_{k=n}(X_1, \dots, X_n)$ . When  $w_i = 1$ , the minimum of  $\mathbf{S}_k$  is  $\frac{1}{k} \sum_{i=1}^k X_i$ , due to the property of order statistics. The maximum of  $\mathbf{S}_k$  is  $\frac{1}{k} \sum_{i=1}^k X_{n-i+1}$ . The monotonicity of the order statistics implies the monotonicity of the extrema with respect to  $k$ , i.e., the support of  $\mathbf{S}_k$  shrinks monotonically. For unequal  $w_i$ s, the shrinkage of the support of  $\mathbf{S}_k$  might not be strictly monotonic, but the general trend remains, since all  $LL$ -statistics converge to the same point, as  $k \rightarrow n$ . Therefore, if  $\frac{\sum_{i=1}^n X_i w_i}{\sum_{i=1}^n w_i}$  approaches the population mean when  $n \rightarrow \infty$ , all  $LL$ -statistics based on such consistent kernel function approach the population mean as  $k \rightarrow \infty$ . For example, if  $whl_k = BM_{\nu, \epsilon_k, n=k}$ ,  $\nu \ll \epsilon_k^{-1}$ ,  $\epsilon_k \rightarrow 0$ , such kernel function is consistent. These cases are termed the  $LL$ -mean ( $LLM_{k, \epsilon, \gamma, n}$ ). By substituting the  $WA_{\epsilon_0, \gamma, n}$  for the  $L_{\epsilon_0, \gamma, n}$  in  $LL$ -statistic, the resulting statistic is referred to as the weighted  $L$ -statistic ( $WL_{k, \epsilon, \gamma, n}$ ). The case having a consistent kernel function is termed as the weighted  $L$ -mean ( $WLM_{k, \epsilon, \gamma, n}$ ). The  $w_i = 1$  case of  $WLM_{k, \epsilon, \gamma, n}$  is termed the weighted Hodges–Lehmann mean ( $WHLM_{k=1, \epsilon, \gamma, n}$ ). The  $WHLM_{k=1, \epsilon, \gamma, n}$  is the weighted average. If  $k \geq 2$  and the  $WA$  in  $WHLM$  is set as  $TM_{\epsilon_0}$ , it is called the trimmed H-L mean (Figure 1,  $k = 2$ ,  $\epsilon_0 = \frac{15}{64}$ ). The  $THLM_{k=2, \epsilon, \gamma=1, n}$  appears similar to the Wilcoxon's one-sample statistic investigated by Saleh in 1976 (41), which involves first censoring the sample, and then computing the

mean of the number of events that the pairwise mean is greater than zero. The THLM $_{k=2, \epsilon=1-(1-\frac{1}{2})^{\frac{1}{2}}, \gamma=1, n}$  is the Hodges-Lehmann estimator, or more generally, a special case of the median Hodges-Lehmann mean ( $mHLM_{k,n}$ ).  $mHLM_{k,n}$  is asymptotically equivalent to the  $MoM_{k,b=\frac{n}{k}}$  as discussed previously. Therefore, it is possible to define a series of location estimators, analogous to the WHLM, based on MoM. For example, the  $\gamma$ -median of means,  $\gamma moM_{k,b=\frac{n}{k}, n}$ , is defined by replacing the median in  $MoM_{k,b=\frac{n}{k}, n}$  with the  $\gamma$ -median.

The  $hl_k$  kernel distribution, denoted as  $F_{hl_k}$ , can be defined as the probability distribution of the sorted sequence,  $\text{sort}\left((hl_k(X_{N_1}, \dots, X_{N_k}))_{N=1}^{\binom{n}{k}}\right)$ . For any real value  $y$ , the cdf of the  $hl_k$  kernel distribution is given by:  $F_{hl_k}(y) = \mathbb{P}(Y_i \leq y)$ , where  $Y_i$  represents an individual element from the sorted sequence. The overall  $hl_k$  kernel distributions possess a two-dimensional structure, encompassing  $n$  kernel distributions with varying  $k$  values, from 1 to  $n$ , where one dimension is inherent to each individual kernel distribution, while the other is formed by the alignment of the same percentiles across all kernel distributions. As  $k$  increases, all percentiles converge to  $\bar{X}$ , leading to the concept of  $\gamma$ - $U$ -orderliness:

$$\begin{aligned} (\forall k_2 \geq k_1 \geq 1, \gamma mHLM_{k_2, \epsilon=1-(\frac{\gamma}{1+\gamma})^{\frac{1}{k_2}}, \gamma} &\geq \gamma mHLM_{k_1, \epsilon=1-(\frac{\gamma}{1+\gamma})^{\frac{1}{k_1}}, \gamma}) \vee \\ (\forall k_2 \geq k_1 \geq 1, \gamma mHLM_{k_2, \epsilon=1-(\frac{\gamma}{1+\gamma})^{\frac{1}{k_2}}, \gamma} &\leq \gamma mHLM_{k_1, \epsilon=1-(\frac{\gamma}{1+\gamma})^{\frac{1}{k_1}}, \gamma}) \end{aligned}$$

where  $\gamma mHLM_k$  sets the WA in WHLM as  $\gamma$ -median, with  $\gamma$  being constant. The direction of the inequality depends on the relative magnitudes of  $\gamma mHLM_{k=1, \epsilon, \gamma} = \gamma m$  and  $\gamma mHLM_{k=\infty, \epsilon, \gamma} = \mu$ . The Hodges-Lehmann inequality can be defined as a special case of the  $\gamma$ - $U$ -orderliness when  $\gamma = 1$ . When  $\gamma \in \{0, \infty\}$ , the  $\gamma$ - $U$ -orderliness is valid for any distribution as previously shown. If  $\gamma \notin \{0, \infty\}$ , analytically proving the validity of the  $\gamma$ - $U$ -orderliness for a parametric distribution is pretty challenging. As an example, the  $hl_2$  kernel distribution has a probability density function  $f_{hl_2}(x) = \int_0^{2x} 2f(t)f(2x-t)dt$  (a result after the transformation of variables); the support of the original distribution is assumed to be  $[0, \infty)$  for simplicity. The expected value of the H-L estimator is the positive solution of  $\int_0^{H-L} (f_{hl_2}(s))ds = \frac{1}{2}$ . For the exponential distribution,  $f_{hl_2, exp}(x) = 4\lambda^{-2}xe^{-2\lambda^{-1}x}$ ,  $\lambda$  is a scale parameter,  $E[H-L] = \frac{-W_{-1}(-\frac{1}{2e})-1}{2}\lambda \approx 0.839\lambda$ , where  $W_{-1}$  is a branch of the Lambert  $W$  function which cannot be expressed in terms of elementary functions. However, the violation of the  $\gamma$ - $U$ -orderliness is bounded under certain assumptions, as shown below.

**Theorem .17.** For any distribution with a finite second central moment,  $\sigma^2$ , the following concentration bound can be established for the  $\gamma$ -median of means,

$$\mathbb{P}\left(\gamma moM_{k,b=\frac{n}{k}, n} - \mu > \frac{t\sigma}{\sqrt{k}}\right) \leq e^{-\frac{2n}{k}\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)^2}.$$

*Proof.* Denote the mean of each block as  $\hat{\mu}_i$ ,  $1 \leq i \leq b$ . Observe that the event  $\left\{\gamma moM_{k,b=\frac{n}{k}, n} - \mu > \frac{t\sigma}{\sqrt{k}}\right\}$  necessitates the condition that there are at least  $b(1 - \frac{\gamma}{1+\gamma})$  of  $\hat{\mu}_i$ s larger than  $\mu$  by more than  $\frac{t\sigma}{\sqrt{k}}$ , i.e.,  $\left\{\gamma moM_{k,b=\frac{n}{k}, n} - \mu > \frac{t\sigma}{\sqrt{k}}\right\} \subset$

$\left\{\sum_{i=1}^b \mathbf{1}_{(\hat{\mu}_i - \mu) > \frac{t\sigma}{\sqrt{k}}} \geq b(1 - \frac{\gamma}{1+\gamma})\right\}$ , where  $\mathbf{1}_A$  is the indicator of event  $A$ . Assuming a finite second central moment,  $\sigma^2$ , it follows from one-sided Chebeshev's inequality that  $\mathbb{E}\left(\mathbf{1}_{(\hat{\mu}_i - \mu) > \frac{t\sigma}{\sqrt{k}}}\right) = \mathbb{P}\left((\hat{\mu}_i - \mu) > \frac{t\sigma}{\sqrt{k}}\right) \leq \frac{\sigma^2}{k\sigma^2 + t^2\sigma^2}$ .

Given that  $\mathbf{1}_{(\hat{\mu}_i - \mu) > \frac{t\sigma}{\sqrt{k}}} \in [0, 1]$  are independent and identically distributed random variables, according to the aforementioned inclusion relation, the one-sided Chebeshev's inequality and the one-sided Hoeffding's inequality,  $\mathbb{P}\left(\gamma moM_{k,b=\frac{n}{k}, n} - \mu > \frac{t\sigma}{\sqrt{k}}\right) \leq \mathbb{P}\left(\sum_{i=1}^b \mathbf{1}_{(\hat{\mu}_i - \mu) > \frac{t\sigma}{\sqrt{k}}} \geq b(1 - \frac{\gamma}{1+\gamma})\right) = \mathbb{P}\left(\frac{1}{b} \sum_{i=1}^b \left(\mathbf{1}_{(\hat{\mu}_i - \mu) > \frac{t\sigma}{\sqrt{k}}} - \mathbb{E}\left(\mathbf{1}_{(\hat{\mu}_i - \mu) > \frac{t\sigma}{\sqrt{k}}}\right)\right) \geq \left(1 - \frac{\gamma}{1+\gamma}\right) - \mathbb{E}\left(\mathbf{1}_{(\hat{\mu}_i - \mu) > \frac{t\sigma}{\sqrt{k}}}\right)\right) \leq e^{-2b\left(\left(1 - \frac{\gamma}{1+\gamma}\right) - \mathbb{E}\left(\mathbf{1}_{(\hat{\mu}_i - \mu) > \frac{t\sigma}{\sqrt{k}}}\right)\right)^2} = e^{-2b\left(1 - \frac{\gamma}{1+\gamma} - \frac{\sigma^2}{k\sigma^2 + t^2\sigma^2}\right)^2} = e^{-2b\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)^2}.$

**Theorem .18.** Let  $B(k, \gamma, t, n) = e^{-\frac{2n}{k}\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)^2}$ . If  $\frac{1}{k} \in \mathbb{N}$ ,  $\gamma \geq 0$ ,  $0 \leq t^2 \leq \gamma + 1$ , and  $\gamma - t^2 + 1 \leq k \leq \frac{1}{2}\sqrt{9\gamma^2 + 18\gamma - 8\gamma t^2 - 8t^2 + 9} + \frac{1}{2}(3\gamma - 2t^2 + 3)$ ,  $B$  is monotonic decreasing with respect to  $k$ .

*Proof.* Since  $\frac{\partial B}{\partial k} = \left(\frac{2n\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)^2}{k^2} - \frac{4n\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)}{k(k+t^2)^2}\right) e^{-\frac{2n\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)^2}{k}}$  and  $n \in \mathbb{N}$ ,  $\frac{\partial B}{\partial k} \leq 0 \Leftrightarrow \frac{2n\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)^2}{k^2} - \frac{4n\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)}{k(k+t^2)^2} \leq 0 \Leftrightarrow \frac{2n(-\gamma + k + t^2 - 1)(k^2 - 3(\gamma + 1)k + 2kt^2 + t^2(-\gamma + t^2 - 1))}{(\gamma + 1)^2 k^2 (k + t^2)^3} \leq 0 \Leftrightarrow (-\gamma + k + t^2 - 1)(k^2 - 3(\gamma + 1)k + 2kt^2 + t^2(-\gamma + t^2 - 1)) \leq 0$ . When the factors are expanded, it yields a cubic inequality in terms of  $k$ :  $k^3 + k^2(3t^2 - 4(\gamma + 1)) + 3k(\gamma - t^2 + 1)^2 + t^2(\gamma - t^2 + 1)^2 \leq 0$ . Assuming  $0 \leq t^2 < \gamma + 1$  and  $\gamma \geq 0$ , using the factored form and subsequently applying the quadratic formula, the inequality is valid if  $\gamma - t^2 + 1 \leq k \leq \frac{1}{2}\sqrt{9\gamma^2 + 18\gamma - 8\gamma t^2 - 8t^2 + 9} + \frac{1}{2}(3\gamma - 2t^2 + 3)$ .  $\square$

Let  $X$  be a random variable and  $\bar{Y} = \frac{1}{k}(Y_1 + \dots + Y_k)$  be the average of  $k$  independent, identically distributed copies of  $X$ . Applying the variance operation gives:  $\text{Var}(\bar{Y}) = \text{Var}\left(\frac{1}{k}(Y_1 + \dots + Y_k)\right) = \frac{1}{k^2}(\text{Var}(Y_1) + \dots + \text{Var}(Y_k)) = \frac{1}{k^2}(k\sigma^2) = \frac{\sigma^2}{k}$ , since the variance operation is a linear operator for independent variables, and the variance of a scaled random variable is the square of the scale times the variance of the variable, i.e.,  $\text{Var}(cX) = E[(cX - E[cX])^2] = E[(cX - cE[X])^2] = E[c^2(X - E[X])^2] = c^2E[(X - E[X])^2] = c^2\text{Var}(X)$ . Thus, the standard deviation of the  $hl_k$  kernel distribution, asymptotically, is  $\frac{\sigma}{\sqrt{k}}$ . By utilizing the asymptotic bias bound of any quantile for any continuous distribution with a finite second central moment,  $\sigma^2$  (2), a conservative asymptotic bias bound of  $\gamma moM_{k,b=\frac{n}{k}}$  can be estab-

lished as  $\gamma \text{MoM}_{k,b=\frac{n}{k}} - \mu \leq \sqrt{\frac{\gamma}{1-\frac{\gamma}{1+\gamma}}} \sigma_{hl_k} = \sqrt{\frac{\gamma}{k}} \sigma$ . That implies in Theorem .17,  $t < \sqrt{\gamma}$ , so when  $\gamma = 1$ , the upper bound of  $k$ , subject to the monotonic decreasing constraint, is  $2 + \sqrt{5} < \frac{1}{2} \sqrt{9 + 18 - 8t^2 - 8t^2 + 9} + \frac{1}{2} (3 - 2t^2 + 3) \leq 6$ , the lower bound is  $1 < 2 - t^2 \leq 2$ . These analyses elucidate a surprising result: although the conservative asymptotic bound of  $\text{MoM}_{k,b=\frac{n}{k}}$  is monotonic with respect to  $k$ , its concentration bound is optimal when  $k \in (2 + \sqrt{5}, 6]$ .

Then consider the structure within each individual  $hl_k$  kernel distribution. The sorted sequence  $\mathbf{S}_k$ , when  $k = n - 1$ , has  $n$  elements and the corresponding  $hl_k$  kernel distribution can be seen as a location-scale transformation of the original distribution, so the corresponding  $hl_k$  kernel distribution is  $\nu$ th  $\gamma$ -ordered if and only if the original distribution is  $\nu$ th  $\gamma$ -ordered according to Theorem .2. Analytically proving other cases is challenging. For example,  $f'_{hl_2}(x) = 4f(2x)f(0) + \int_0^{2x} 4f(t)f'(2x-t)dt$ , the strict negative of  $f'_{hl_2}(x)$  is not guaranteed if just assuming  $f'(x) < 0$ , so, even if the original distribution is monotonic decreasing, the  $hl_2$  kernel distribution might be non-monotonic. Also, unlike the pairwise difference distribution, if the original distribution is unimodal, the pairwise mean distribution might be non-unimodal, as demonstrated by a counterexample given by Chung in 1953 and mentioned by Hodges and Lehmann in 1954 (42, 43). Theorem .9 implies that the violation of  $\nu$ th  $\gamma$ -orderliness within the  $hl_k$  kernel distribution is also bounded, and the bound monotonically shrinks as  $k$  increases because the bound is in unit of the standard deviation of the  $hl_k$  kernel distribution. If all  $hl_k$  kernel distributions are  $\nu$ th  $\gamma$ -ordered and the distribution itself is  $\nu$ th  $\gamma$ -ordered and  $\gamma$ - $U$ -ordered, then the distribution is called  $\nu$ th  $\gamma$ - $U$ -ordered. The following theorem highlights the significance of symmetric distribution.

**Theorem .19.** *Any symmetric distribution is  $\nu$ th  $U$ -ordered.*

*Proof.* A random variable is symmetric about zero if and only if its characteristic function is real valued. Since the characteristic function of the average of  $k$  independent, identically distributed random variables is the product of the  $k$ th root of their individual characteristic functions:  $\varphi_{\bar{Y}}(t) = \prod_{r=1}^k (\varphi_{Y_r}(t))^{\frac{1}{k}}$ ,  $\bar{Y}$  is symmetric. The conclusion follows immediately from the definition of  $\nu$ th  $U$ -orderliness and Theorem .2, .3, and .4.  $\square$

The succeeding theorem shows that the  $whl_k$  kernel distribution is invariably a location-scale distribution if the original distribution belongs to a location-scale family with the same location and scale parameters.

**Theorem .20.**  $whl_k(x_1 = \lambda x_1 + \mu, \dots, x_k = \lambda x_k + \mu) = \lambda whl_k(x_1, \dots, x_k) + \mu$ .

*Proof.*  $whl_k(x_1 = \lambda x_1 + \mu, \dots, x_k = \lambda x_k + \mu) = \frac{\sum_{i=1}^k (\lambda x_i + \mu) w_i}{\sum_{i=1}^k w_i} = \frac{\sum_{i=1}^k \lambda x_i w_i + \sum_{i=1}^k \mu w_i}{\sum_{i=1}^k w_i} = \lambda \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i} + \frac{\sum_{i=1}^k \mu w_i}{\sum_{i=1}^k w_i} = \lambda \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i} + \mu = \lambda whl_k(x_1, \dots, x_k) + \mu$ .  $\square$

According to Theorem .20, the  $\gamma$ -weighted inequality for a right-skewed distribution can be modified as  $\forall 0 \leq \epsilon_{01} \leq \epsilon_{02} \leq \frac{1}{1+\gamma}$ ,  $\text{WLM}_{k,\epsilon=1-(1-\epsilon_{01})^{\frac{1}{k},\gamma}} \geq \text{WLM}_{k,\epsilon=1-(1-\epsilon_{02})^{\frac{1}{k},\gamma}}$ ,

which holds the same rationale as the  $\gamma$ -weighted inequality defined in the last section. If the  $\nu$ th  $\gamma$ -orderliness is valid for the  $whl_k$  kernel distribution, then all results in the last section can be directly implemented. From that, the binomial H-L mean (set the WA as BM) can be constructed (Figure 1), while its maximum breakdown point is  $\approx 0.065$  if  $\nu = 3$ . A comparison of the biases of  $\text{STM}_{\epsilon=\frac{1}{8}}$ ,  $\text{SWM}_{\epsilon=\frac{1}{8}}$ ,  $\text{BWM}_{\epsilon=\frac{1}{8}}$ ,  $\text{BM}_{\nu=2,\epsilon=\frac{1}{8}}$ ,  $\text{BM}_{\nu=3,\epsilon=\frac{1}{8}}$ ,  $\text{SQM}_{\epsilon=\frac{1}{8}}$ ,  $\text{THLM}_{k=2,\epsilon=\frac{1}{8}}$ ,  $\text{WiHLM}_{k=2,\epsilon=\frac{1}{8}}$  (Winsorized H-L mean),  $\text{SQHLM}_{k=\frac{2\ln(2)-\ln(3)}{3\ln(2)-\ln(7)},\epsilon=\frac{1}{8}}$ ,  $\text{mHLM}_{k=\frac{\ln(2)}{3\ln(2)-\ln(7)},\epsilon=\frac{1}{8}}$ ,  $\text{THLM}_{k=5,\epsilon=\frac{1}{8}}$ , and  $\text{WiHLM}_{k=5,\epsilon=\frac{1}{8}}$  is appropriate (Figure 1, SI Dataset S1), given their same breakdown points, with  $\text{mHLM}_{k=\frac{\ln(2)}{3\ln(2)-\ln(7)},\epsilon=\frac{1}{8}}$  exhibiting the smallest biases. Another comparison among the H-L estimator, the trimmed mean, and the Winsorized mean, all with the same breakdown point, yields the same result that the H-L estimator has the smallest biases (SI Dataset S1). This aligns with Devroye et al. (2016) and Laforgue, Clemencon, and Bertail (2019)'s seminal works that  $\text{MoM}_{k,b=\frac{n}{k}}$  and  $\text{MoRM}_{k,b,n}$  are nearly optimal with regards to concentration bounds for heavy-tailed distributions (18, 19).

In 1958, Richtmyer introduced the concept of quasi-Monte Carlo simulation that utilizes low-discrepancy sequences, resulting in a significant reduction in computational expenses for large sample simulation (44). Among various low-discrepancy sequences, Sobol sequences are often favored in quasi-Monte Carlo methods (45). Building upon this principle, in 1991, Do and Hall extended it to bootstrap and found that the quasi-random approach resulted in lower variance compared to other bootstrap Monte Carlo procedures (46). By using a deterministic approach, the variance of  $\text{mHLM}_{k,n}$  is much lower than that of  $\text{MoM}_{k,b=\frac{n}{k}}$  (SI Dataset S1), when  $k$  is small. This highlights the superiority of the median Hodges-Lehmann mean over the median of means, as it not only can provide an accurate estimate for moderate sample sizes, but also allows the use of quasi-bootstrap, where the bootstrap size can be adjusted as needed.

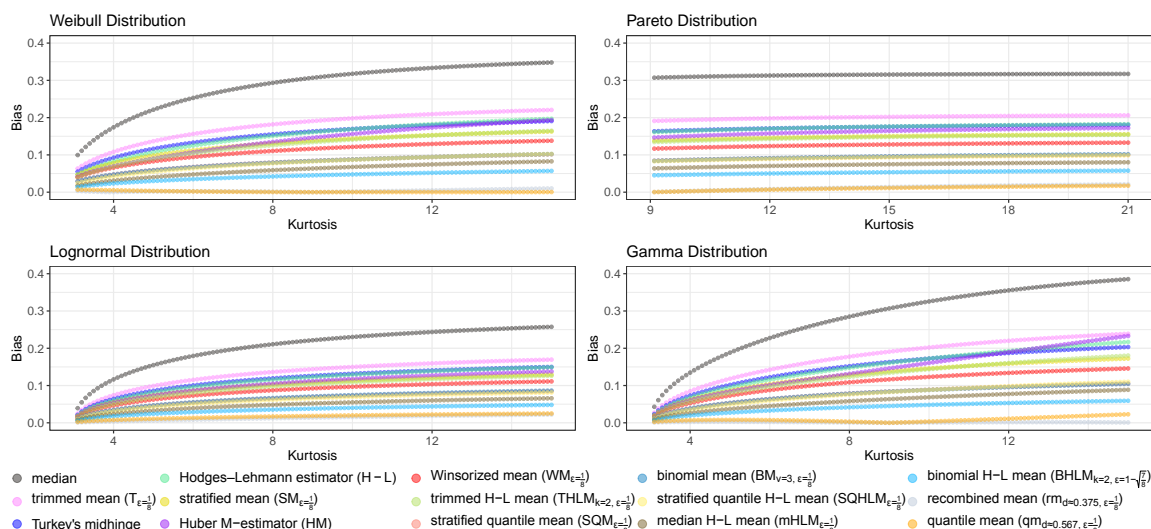
## Methods

The robust location estimates presented in Figure 1 and SI Dataset S1 were obtained using large quasi-random samples (44, 45) with sample size 3.686 million for the Weibull, gamma, Pareto, and lognormal distributions within specified kurtosis ranges as shown in Figure 1 to study the large sample performance. The standard errors of these estimators were computed by approximating the sampling distribution using 1000 pseudorandom samples of size  $n = 5184$  for these distribution and the generalized Gaussian distributions with the parameter settings detailed in the SI Text.

**Data and Software Availability.** Data for Figure 1 are given in SI Dataset S1. All codes have been deposited in [GitHub](#).

1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae*. (Henricus Dieterich), (1823).
2. 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).
3. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
4. P Daniell, Observations weighted according to order. *Am. J. Math.* **42**, 222–236 (1920).
5. JW Tukey, A survey of sampling from contaminated distributions in *Contributions to probability and statistics*. (Stanford University Press), pp. 448–485 (1960).
6. WJ Dixon, Simplified Estimation from Censored Normal Samples. *The Annals Math. Stat.* **31**, 385–391 (1960).





**Fig. 1.** Standardized biases (with respect to  $\mu$ ) of fifteen robust location estimates (including two parametric estimators from REDS II for better comparison) on large quasi-random samples in four two-parameter right skewed unimodal distributions, as a function of the kurtosis. The methods are described in the SI Text.

7. K Danielak, T Rychlik, Theory & methods: Exact bounds for the bias of trimmed means. *Aust. & New Zealand J. Stat.* **45**, 83–96 (2003).
8. M Bieniek, Comparison of the bias of trimmed and winsorized means. *Commun. Stat. Methods* **45**, 6641–6650 (2016).
9. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
10. F Wilcoxon, Individual comparisons by ranking methods. *Biom. Bull.* **1**, 80–83 (1945).
11. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
12. Q Sun, WX Zhou, J Fan, Adaptive huber regression. *J. Am. Stat. Assoc.* **115**, 254–265 (2020).
13. O Catoni, Challenging the empirical mean and empirical variance: a deviation study. *Annales de l'IHP Probab. et statistiques* **48**, 1148–1185 (2012).
14. AS Nemirovskij, DB Yudin, *Problem complexity and method efficiency in optimization.* (Wiley-Interscience), (1983).
15. MR Jerrum, LG Valiant, VV Vazirani, Random generation of combinatorial structures from a uniform distribution. *Theor. computer science* **43**, 169–188 (1986).
16. 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).
17. D Hsu, S Sabato, Heavy-tailed regression with a generalized median-of-means in *International Conference on Machine Learning.* (PMLR), pp. 37–45 (2014).
18. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.* **44**, 2695–2725 (2016).
19. P Laforgue, S Cléménçon, P Bertail, On medians of (randomized) pairwise means in *International Conference on Machine Learning.* (PMLR), pp. 1272–1281 (2019).
20. G LECUÉ, M LERASLE, Robust machine learning by median-of-means: Theory and practice. *The Annals Stat.* **48**, 906–931 (2020).
21. B Efron, Bootstrap methods: Another look at the jackknife. *The Annals Stat.* **7**, 1–26 (1979).
22. PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. *The annals statistics* **9**, 1196–1217 (1981).
23. PJ Bickel, DA Freedman, Asymptotic normality and the bootstrap in stratified sampling. *The annals statistics* **12**, 470–482 (1984).
24. R Helmers, P Janssen, N Veraverbeke, *Bootstrapping U-quantiles.* (CWI. Department of Operations Research, Statistics, and System Theory [BS]), (1990).
25. 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).
26. G McIntyre, A method for unbiased selective sampling, using ranked sets. *Aust. journal agricultural research* **3**, 385–390 (1952).
27. PL Davies, U Gather, Breakdown and groups. *The Annals Stat.* **33**, 977 – 1035 (2005).
28. DL Donoho, PJ Huber, The notion of breakdown point. *A festschrift for Erich L. Lehmann* **157184** (1983).
29. 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).
30. PJ Bickel, On adaptive estimation. *The Annals Stat.* **10**, 647–671 (1982).
31. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparametric models.* (Springer) Vol. 4, (1993).
32. JT Runnenburg, Mean, median, mode. *Stat. Neerlandica* **32**, 73–79 (1978).
33. Wv Zwet, Mean, median, mode ii. *Stat. Neerlandica* **33**, 1–5 (1979).
34. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
35. 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).
36. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
37. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).
38. JW Tukey, *Exploratory data analysis.* (Reading, MA) Vol. 2, (1977).
39. PK Sen, On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. *Biometrics* **19**, 532–552 (1963).
40. RJ Serfling, Generalized L-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).
41. A Ehsanes Saleh, Hodges-lehmann estimate of the location parameter in censored samples. *Annals Inst. Stat. Math.* **28**, 235–247 (1976).
42. J Hodges, E Lehmann, Matching in paired comparisons. *The Annals Math. Stat.* **25**, 787–791 (1954).
43. K Chung, Sur les lois de probabilité unimodales. *COMPTEs RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES* **236**, 583–584 (1953).
44. RD Richtmyer, A non-random sampling method, based on congruences, for " monte carlo" problems, (New York Univ., New York. Atomic Energy Commission Computing and Applied . . .), Technical report (1958).
45. IM Sobol', On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **7**, 784–802 (1967).
46. KA Do, P Hall, Quasi-random resampling for the bootstrap. *Stat. Comput.* **1**, 13–22 (1991).