Robust estimations for semiparametric models: Mean

Tuban Lee

This manuscript was compiled on July 9, 2023

As one of the most fundamental problems in statistics, robust location estimation has many prominent solutions, such as the symmetric trimmed mean, symmetric Winsorized mean, Hodges–Lehmann estimator, Huber M-estimator, and median of means. Recent studies suggest that their biases concerning the mean can be quite different in asymmetric distributions, but the underlying mechanisms largely remain unclear. This study exploited a semiparametric method to classify distributions by the asymptotic orderliness of location estimates 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. Building on the γ-U-orderliness, the superiority of the median Hodges–Lehmann mean is discussed.

In 1823, Gauss (1) proved that for any unimodal distribution, $|m - \mu| \leq \sqrt{\frac{3}{4}}\omega$ and $\sigma \leq \omega \leq 2\sigma$, where μ is the population mean, m is the population median, ω is the root mean square deviation from the mode, and σ is the population standard deviation. This pioneering work revealed that, the potential bias of the median, the most fundamental robust location estimate, with respect to the mean is bounded in units of a scale parameter under certain assumptions. Bernard, Kazzi, and Vanduffel (2020) (2) further derived asymptotic bias bounds for any quantile in unimodal distributions with finite second moments. They showed that m has the smallest maximum distance to μ among all symmetric quantile averages (SQA_ϵ). Daniell, in 1920, (3) analyzed a class of estimators, linear combinations of order statistics, and identified that the ϵ-symmetric trimmed mean (STM_ϵ) belongs to this class. Another popular choice, the ϵ-symmetric Winsorized mean (SWM_ϵ), named after Winsor and introduced by Tukey (4) and Dixon (5) in 1960, is also an L-estimator. Bieniek (2016) derived exact bias upper bounds of the Winsorized mean based on Danielak and Rychlik's work (2003) on the trimmed mean for any distribution with a finite second moment and confirmed that the former is smaller than the latter (6, 7). In 1963, Hodges and Lehmann (8) proposed a class of nonparametric location estimators based on rank tests and, from the Wilcoxon signed-rank statistic (9), deduced the median of pairwise means as a robust location estimator for a symmetric population. Both L-statistics and R-statistics achieve robustness essentially by removing a certain proportion of extreme values. In 1964, Huber (10) generalized maximum likelihood estimation to the minimization of the sum of a specific loss function, which measures the residuals between the data points and the model's parameters. Some L-estimators are also M-estimators, e.g., the sample mean is an M-estimator with a squared error loss function, the sample median is an M-estimator with an absolute error loss function (10). The Huber M-estimator is obtained by applying the Huber loss function that combines elements of both squared error and absolute error to achieve robustness against gross errors and high efficiency for contaminated Gaussian distributions (10). Sun, Zhou, and Fan (2020) examined the concentration bounds of the Huber M-estimator (11). Mathieu (2022) (12) further derived the concentration bounds of M-estimators and demonstrated that, by selecting the tuning parameter which depends on the variance, the Huber M-estimator can also be a sub-Gaussian estimator. The concept of the median of means ($\text{MoM}_{k,b=\frac{n}{k},n}$) was first introduced by Nemirovsky and Yudin (1983) in their work on stochastic optimization (13). Given its good performance even for distributions with infinite second moments, the MoM has received increasing attention over the past decade (14–17). Devroye, Lerasle, Lugosi, and Oliveira (2016) showed that $\text{MoM}_{k,b=\frac{n}{k},n}$ nears the optimum of sub-Gaussian mean estimation with regards to concentration bounds when the distribution has a heavy tail (15). Laforgue, Clemencon, and Bertail (2019) proposed the median of randomized means ($\text{MoRM}_{k,b,n}$) (16), wherein, rather than partitioning, an arbitrary number, b, of blocks are built independently from the sample, and showed that $\text{MoRM}_{k,b,n}$ has a better non-asymptotic sub-Gaussian property compared to $\text{MoM}_{k,b=\frac{n}{k},n}$. In fact, asymptotically, the Hodges-Lehmann (H-L) estimator is equivalent to $\text{MoM}_{k=2,b=\frac{n}{k}}$ and $\text{MoRM}_{k=2,b}$, 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) (18), Bickel and Freedman (1981, 1984) (19, 20), and Helmers, Janssen, and Veraverbeke (1990) (21).

Significance Statement

In 1964, van Zwet introduced the convex transformation order for comparing the skewness of two distributions. This paradigm shift played a fundamental role in defining robust measures of distributions, from spread to kurtosis. Here, instead of examining the stochastic ordering between two distributions, the orderliness of quantile averages within a distribution is investigated. By classifying distributions through the signs of derivatives, 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.

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

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

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

where $X_1 \leq \ldots \leq X_n$ denote the order statistics of a sample of n independent and identically distributed random variables $X_1, \ldots, X_n$. $b \in \mathbb{N}$, $b \geq 3$. The definition was further refined to guarantee the continuity of the breakdown point by incorporating an additional block in the center when $\lfloor \frac{b-1}{2b\epsilon} \rfloor \bmod 2 = 0$, or by adjusting the central block when $\lfloor \frac{b-1}{2b\epsilon} \rfloor \bmod 2 = 1$ (SI Text). If the subscript n is omitted, only the asymptotic behavior is considered. If b is omitted, $b = 3$ is assumed. $\mathrm{SM}_{\epsilon,b=3}$ is equivalent to STM_{ϵ}, when $\epsilon > \frac{1}{6}$. When $\frac{b-1}{2\epsilon} \in \mathbb{N}$ and $b \bmod 2 = 1$, the basic idea of the stratified mean is to distribute the data into $\frac{b-1}{2\epsilon}$ equal-sized non-overlapping blocks according to their order. Then, further sequentially group these blocks into b equal-sized strata and compute the mean of the middle stratum, which is the median of means of each stratum. In situations where $i \bmod 1 \neq 0$, a potential solution is to generate multiple smaller samples that satisfy the equality by sampling without replacement, and subsequently calculate the mean of all estimations. The details of determining the 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, $\mathrm{SM}_{\epsilon,b,n}$ is different from $\mathrm{MoM}_{k=\frac{n}{b},b,n}$ 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) (22) or ranked set sampling (23), introduced by McIntyre in 1952, as these sampling methods aim to obtain more representative samples or improve the efficiency of sample estimates, but the sample means based on them are not robust. When $b \bmod 2 = 1$, the stratified mean can be regarded as replacing the other equal-sized strata with the middle stratum, which, in principle, is analogous to the Winsorized mean that replaces extreme values with less extreme percentiles. Furthermore, while the bounds confirm that the Winsorized mean and median of means outperform the trimmed mean (6, 7, 15) 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 semiparametric models using the signs of derivatives, reveal their elegant interrelations and connections to parametric models, and show that by exploiting these models, a set of sophisticated mean estimators can be deduced, which exhibit strong robustness to departures from assumptions.

Quantile Average and Weighted Average

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

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

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

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

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

and the second definition is:

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

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

Analogously, the weighted average can be defined as

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

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

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

Definition .1 (Upper/lower breakdown point). The upper breakdown point is the breakdown point generalized in Davies and Gather (2005)'s paper (**?**). The finite-sample upper breakdown point is the finite sample breakdown point defined by Donoho and Huber (1983) (24) and also detailed in (**?**). 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 (25) and proposed a necessary condition for this type of problem, a contribution later explicitly recognized as initiating the field of semiparametric statistics (26). In 1982, Bickel simplified Stein's general heuristic necessary condition (25), derived sufficient conditions, and used them in formulating adaptive estimates (26). 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 (27), 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) (28, 29) endeavored to determine sufficient conditions for the mean-median-mode inequality to hold, thereby implying the possibility of its violation. The class of unimodal distributions that satisfy the mean-median-mode inequality constitutes a subclass of $\mathcal{P}_U$, denoted by $\mathcal{P}_{MMM} \subsetneq \mathcal{P}_U$. To further investigate the relations of location estimates within a distribution, the γ-orderliness for a right-skewed distribution is defined as

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

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

Theorem .1. *A distribution is γ-ordered if and only if its pdf satisfies the inequality $f(Q(\gamma\epsilon)) \geq f(Q(1-\epsilon))$ for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$ or $f(Q(\gamma\epsilon)) \leq f(Q(1-\epsilon))$ for all $0 \leq \epsilon \leq \frac{1}{1+\gamma}$.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Theorem .7 also reveals a relation between convex transformation and orderliness, since ϕ is the non-decreasing convex function in van Zwet's trailblazing work *Convex transformations of random variables* (30) if adding an additional constraint that $\phi'\left(x\right)\geq 0$. Consider a near-symmetric distribution S, such that the $\mathrm{SQA}(\epsilon)$ as a function of ϵ fluctuates from 0 to $\frac{1}{2}$. By definition, S is not ordered. Let s be the pdf of S. Applying the transformation $\phi\left(x\right)$ to S decreases $s(Q_S(\epsilon))$, and the decrease rate, due to the order, is much smaller for $s(Q_S(1-\epsilon))$. As a consequence, as $\phi^{(2)}\left(x\right)$ increases, eventually, after a point, for all $0\leq\epsilon\leq\frac{1}{1+\gamma}$, $s(Q_S(\epsilon))$ becomes greater than $s(Q_S(1-\epsilon))$ even if it was not previously. Thus, the $\mathrm{SQA}(\epsilon)$ function becomes monotonically decreasing, and S becomes ordered. Accordingly, in a family of distributions that differ by a skewness-increasing transformation in van Zwet's sense, violations of orderliness typically occur only when the distribution is near-symmetric.

Pearson proposed using the 3 times standardized mean-median difference, $\frac{3(\mu-m)}{\sigma}$, as a measure of skewness in 1895 (31). Bowley (1926) proposed a measure of skewness based on the $\mathrm{SQA}_{\epsilon=\frac{1}{4}}$-median difference $\mathrm{SQA}_{\epsilon=\frac{1}{4}}-m$ (32). Groeneveld and Meeden (1984) (33) generalized these measures of skewness based on van Zwet's convex transformation (30) while exploring their properties. A distribution is called monotonically right-skewed if and only if $\forall 0\leq\epsilon_1\leq\epsilon_2\leq\frac{1}{2}, \mathrm{SQA}_{\epsilon_1}-m\geq\mathrm{SQA}_{\epsilon_2}-m$. Since m is a constant, the monotonic skewness is equivalent to the orderliness. For a nonordered distribution, the signs of $\mathrm{SQA}_\epsilon - m$ with different breakdown points might be different, implying that some skewness measures indicate left-skewed distribution, while others suggest right-skewed distribution. Although it seems reasonable that such a distribution is likely be generally near-symmetric, counterexamples can be constructed. For example, first consider the Weibull distribution, when $\alpha>\frac{1}{1-\ln(2)}$, it is near-symmetric and nonordered, the non-monotonicity of the SQA function arises when ϵ is close to $\frac{1}{2}$, but if then replacing the third quartile with one from a right-skewed heavy-tailed distribution leads to a right-skewed, heavy-tailed, and nonordered distribution. Therefore, the validity of robust measures of skewness based on the SQA-median difference is closely related to the orderliness of the distribution.

Remarkably, in 2018, Li, Shao, Wang, Yang (34) proved the bias bound of any quantile for arbitrary continuous distributions with finite second moments. Here, let $\mathcal{P}_{\mu,\sigma}$ denotes the set of continuous distributions whose mean is μ and standard deviation is σ. The bias upper bound of the quantile average for $P\in\mathcal{P}_{\mu=0,\sigma=1}$ is given in the following theorem.

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

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

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

Proof. Since $\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}}\frac{1}{2}(Q(\gamma\epsilon)+Q(1-\epsilon))\leq\frac{1}{2}(\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}}Q(\gamma\epsilon)+\sup_{P\in\mathcal{P}_{\mu=0,\sigma=1}}Q(1-\epsilon))$, the assertion follows directly from the Lemma 2.6 in (34). □

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

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

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

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

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

Proof. When $0\leq\epsilon\leq\frac{1}{6}$, $\frac{\partial\sup\mathrm{QA}}{\partial\epsilon}=\frac{\gamma}{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2}-\frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2}=\frac{\sqrt{\gamma}}{\sqrt{\frac{\epsilon}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2}-\frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2}$. If $\gamma=0$ and $\epsilon\to 0^+$, $\frac{\partial\sup\mathrm{QA}}{\partial\epsilon}=-\frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2}<0$. If $\epsilon\to 0^+$, $\lim_{\epsilon\to 0^+}\left(\frac{\gamma}{(4-3\gamma\epsilon)^2\sqrt{\frac{\epsilon\gamma}{12-9\gamma\epsilon}}}-\frac{1}{3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2}\right)=\lim_{\epsilon\to 0^+}\left(\frac{\sqrt{3\gamma}}{\sqrt{4^3\epsilon}}-\frac{1}{6\sqrt{\epsilon^3}}\right)\to-\infty$, for all $0\leq\gamma\leq 1$, so, $\frac{\partial\sup\mathrm{QA}}{\partial\epsilon}<0$. When $0<\epsilon\leq\frac{1}{6}$ and $0<\gamma\leq 1$, to prove $\frac{\partial\sup\mathrm{QA}}{\partial\epsilon}\leq 0$, it is equivalent to showing $\frac{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2}{\gamma}\geq 3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2$. Define $L(\epsilon,\gamma)=\frac{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2}{\gamma}$, $R(\epsilon,\gamma)=3\sqrt{\frac{4}{\epsilon}-9}\epsilon^2$. $\frac{L(\epsilon,\gamma)}{\epsilon^2}=\frac{\sqrt{\frac{\epsilon\gamma}{12-9\epsilon\gamma}}(4-3\epsilon\gamma)^2}{\gamma\epsilon^2}=\frac{1}{\gamma}\left(\frac{4}{\epsilon}-3\gamma\right)^2\sqrt{\frac{1}{\frac{12}{\epsilon\gamma}-9}}$, $\frac{R(\epsilon,\gamma)}{\epsilon^2}=3\sqrt{\frac{4}{\epsilon}-9}$. Then, the above inequality is equivalent to $\frac{L(\epsilon,\gamma)}{\epsilon^2}\geq\frac{R(\epsilon,\gamma)}{\epsilon^2}\Leftrightarrow\frac{1}{\gamma}\sqrt{\frac{1}{\frac{12}{\epsilon\gamma}-9}}\left(\frac{4}{\epsilon}-3\gamma\right)^2\geq 3\sqrt{\frac{4}{\epsilon}-9}\Leftrightarrow\frac{1}{\gamma}\left(\frac{4}{\epsilon}-3\gamma\right)^2\geq 3\sqrt{\frac{12}{\epsilon\gamma}-9}\sqrt{\frac{4}{\epsilon}-9}\Leftrightarrow\frac{1}{\gamma^2}\left(\frac{4}{\epsilon}-3\gamma\right)^4\geq 9\left(\frac{12}{\epsilon\gamma}-9\right)\left(\frac{4}{\epsilon}-9\right)$. Let $LmR\left(\frac{1}{\epsilon}\right)=\frac{1}{\gamma^2}\left(\frac{4}{\epsilon}-3\gamma\right)^4-9\left(\frac{12}{\epsilon\gamma}-9\right)\left(\frac{4}{\epsilon}-9\right)$. $\frac{\partial LmR(1/\epsilon)}{\partial(1/\epsilon)}=\frac{16\left(\frac{4}{\epsilon}-3\gamma\right)^3}{\gamma^2}-$

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

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

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

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

Robust Mean Estimators

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

Theorem .11. *For a distribution that is right-skewed and follows the γ-trimming inequality, it is asymptotically true that the quantile average is always greater or equal to the corresponding trimmed mean with the same ϵ and γ, for all $0\leq\epsilon\leq\frac{1}{1+\gamma}$.*

Proof. According to the definition of the γ-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 δ 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 $\delta\to 0^+$, $\frac{1}{2\delta}\left(\int_{1-\epsilon}^{1-\epsilon+\delta}Q(u)\,du+\int_{\gamma\epsilon-\delta}^{\gamma\epsilon}Q(u)\,du\right)=\frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2}\geq\frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du$, the proof is complete. □

An analogous result about the relation between the ϵ,γ-trimmed mean and the ϵ,γ-Winsorized mean can be obtained in the following theorem.

Theorem .12. *For a right-skewed distribution following the γ-trimming inequality, asymptotically, the Winsorized mean is always greater or equal to the corresponding trimmed mean with the same ϵ and γ, for all $0\leq\epsilon\leq\frac{1}{1+\gamma}$, provided that $0\leq\gamma\leq 1$. If assuming γ-orderliness, the inequality is valid for any non-negative γ.*

Proof. According to Theorem .11, $\frac{Q(\gamma\epsilon)+Q(1-\epsilon)}{2}\geq\frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du\Leftrightarrow\gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right)\geq\left(\frac{2\gamma\epsilon}{1-\epsilon-\gamma\epsilon}\right)\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du$. Then, if $0\leq\gamma\leq 1$, $\left(1-\frac{1}{1-\epsilon-\gamma\epsilon}\right)\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du+\gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right)\geq 0\Rightarrow\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du+\gamma\epsilon Q(\gamma\epsilon)+\epsilon Q(1-\epsilon)\geq\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du+\gamma\epsilon\left(Q(\gamma\epsilon)+Q(1-\epsilon)\right)\geq\frac{1}{1-\epsilon-\gamma\epsilon}\int_{\gamma\epsilon}^{1-\epsilon}Q(u)\,du$, the proof of the first assertion is complete. The second assertion is established in Theorem 0.3. in the SI Text. □

Replacing the TM in the γ-trimming inequality with WA forms the definition of the γ-weighted inequality. The γ-orderliness also implies the γ-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 μ and a scale parameter λ, asymptotically, any $\mathrm{WA}(\epsilon, \gamma)$ can be expressed as $\lambda \mathrm{WA}_0(\epsilon, \gamma) + \mu$, where $\mathrm{WA}_0(\epsilon, \gamma)$ 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 γ-weighted inequality is whether the family of probability distributions also adheres to the γ-weighted inequality.

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

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

which is double weighting the leftest and rightest blocks having sizes of $\gamma\epsilon n$ and ϵn, respectively. As a consequence of $\mathcal{B}_i - \mathcal{B}_{i+1} \geq 0$, the γ-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 .13. *Asymptotically, for a right-skewed distribution following the γ-orderliness, the Winsorized mean is always greater than or equal to the corresponding block Winsorized mean with the same ϵ and γ, 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 $\mathrm{WM}l(x) = Q(\gamma\epsilon)$ and $\mathrm{BWM}l(x) = Q(x)$. In both functions, the interval for x is specified as $[\gamma\epsilon, 2\gamma\epsilon]$. Then, define $\mathrm{WM}u(y) = Q(1-\epsilon)$ and $\mathrm{BWM}u(y) = Q(y)$. In both functions, the interval for y is specified as $[1-2\epsilon, 1-\epsilon]$. The function $y : [\gamma\epsilon, 2\gamma\epsilon] \to [1-2\epsilon, 1-\epsilon]$ defined by $y(x) = 1 - \frac{x}{\gamma}$ is a bijection. $\mathrm{WM}l(x) + \mathrm{WM}u(y(x)) = Q(\gamma\epsilon) + Q(1-\epsilon) \geq \mathrm{BWM}l(x) + \mathrm{BWM}u(y(x)) = Q(x) + Q\left(1 - \frac{x}{\gamma}\right)$ is valid for all $x \in [\gamma\epsilon, 2\gamma\epsilon]$, according to the definition of γ-orderliness. Integration of the left side yields, $\int_{\gamma\epsilon}^{2\gamma\epsilon} (\mathrm{WM}l(u) + \mathrm{WM}u(y(u)))\, du = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(\gamma\epsilon)\, du + \int_{y(\gamma\epsilon)}^{y(2\gamma\epsilon)} Q(1-\epsilon)\, du = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(\gamma\epsilon)\, du + \int_{1-2\epsilon}^{1-\epsilon} Q(1-\epsilon)\, du = \gamma\epsilon Q(\gamma\epsilon) + \epsilon Q(1-\epsilon)$, while integration of the right side yields $\int_{\gamma\epsilon}^{2\gamma\epsilon} (\mathrm{BWM}l(x) + \mathrm{BWM}u(y(x)))\, dx = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u)\, du + \int_{\gamma\epsilon}^{2\gamma\epsilon} Q\left(1 - \frac{x}{\gamma}\right) dx = \int_{\gamma\epsilon}^{2\gamma\epsilon} Q(u)\, du + \int_{1-2\epsilon}^{1-\epsilon} Q(u)\, du$, 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.

□

From the second γ-orderliness for a right-skewed distribution, $\frac{\partial^2 \mathrm{QA}}{\partial^2 \epsilon} \geq 0 \Rightarrow \forall 0 \leq a \leq 2a \leq 3a \leq \frac{1}{1+\gamma}, \frac{1}{a}\left(\frac{(\mathrm{QA}(3a,\gamma) - \mathrm{QA}(2a,\gamma))}{a} - \frac{(\mathrm{QA}(2a,\gamma) - \mathrm{QA}(a,\gamma))}{a}\right) \geq 0 \Rightarrow$ if $0 \leq \gamma \leq 1$, $\mathcal{B}_i - 2\mathcal{B}_{i+1} + \mathcal{B}_{i+2} \geq 0$. SM_ϵ 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 νth γ-orderliness for a right-skewed distribution, the recurrence relation of the derivatives naturally produces the alternating binomial coefficients,

$$(-1)^\nu \frac{\partial^\nu \mathrm{QA}}{\partial \epsilon^\nu} \geq 0 \Rightarrow \forall 0 \leq a \leq \ldots \leq (\nu+1)a \leq \frac{1}{1+\gamma},$$

$$\frac{(-1)^\nu}{a} \left(\frac{\frac{\mathrm{QA}(\nu a + a, \gamma)^{\cdot^{\cdot^{\cdot}}}}{a} - \frac{^{\cdot^{\cdot^{\cdot}}}\mathrm{QA}(2a,\gamma)}{a}}{a} - \frac{\frac{\mathrm{QA}(\nu a, \gamma)^{\cdot^{\cdot^{\cdot}}}}{a} - \frac{^{\cdot^{\cdot^{\cdot}}}\mathrm{QA}(a,\gamma)}{a}}{a} \right)$$

$$\geq 0 \Leftrightarrow \frac{(-1)^\nu}{a^\nu} \left(\sum_{j=0}^{\nu} (-1)^j \binom{\nu}{j} \mathrm{QA}\left((\nu - j + 1)\, a, \gamma\right) \right) \geq 0$$

$$\Rightarrow \text{if } 0 \leq \gamma \leq 1, \sum_{j=0}^{\nu} (-1)^j \binom{\nu}{j} \mathcal{B}_{i+j} \geq 0.$$

Based on the νth orderliness, the ϵ,γ-binomial mean is introduced as

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

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 ν is not indicated, it defaults to $\nu = 3$. Since the alternating sum of binomial coefficients equals zero, when $\nu \ll \epsilon^{-1}$ and $\epsilon \to 0$, $\mathrm{BM} \to \mu$. The solutions for the continuity of the breakdown point is the same as that in SM and not repeated here. The equalities $\mathrm{BM}_{\nu=1,\epsilon} = \mathrm{BWM}_\epsilon$ and $\mathrm{BM}_{\nu=2,\epsilon} = \mathrm{SM}_{\epsilon,b=3}$ hold, when $\gamma = 1$ and their respective ϵs are identical. Interestingly, the biases of the $\mathrm{SM}_{\epsilon=\frac{1}{9},b=3}$ and the $\mathrm{WM}_{\epsilon=\frac{1}{9}}$ 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 γ-orderliness, the stratified quantile mean can be defined as

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

$\mathrm{SQM}_{\epsilon=\frac{1}{4}}$ is the Tukey's midhinge (35). In fact, SQM is a subcase of SM when $\gamma = 1$ and $b \to \infty$, 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 .14. *For a right-skewed second γ-ordered distribution, asymptotically, $SQM_{\epsilon,\gamma}$ is always greater or equal to the corresponding $BM_{\nu=2,\epsilon,\gamma}$ with the same ϵ and γ, 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 $\mathrm{SQM}_{\epsilon,\gamma}$ and $\mathrm{BM}_{\nu=2,\epsilon,\gamma}$. The computation of $\mathrm{BM}_{\nu=2,\epsilon,\gamma}$ 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 weighted of one, the sequence indicating the weighted or non-weighted status of each block is: $0, 1, 0, 0, 1, 0, \ldots$. Let this sequence be denoted by $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j)$, its formula is $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j) = \left\lfloor \frac{j \bmod 3}{2} \right\rfloor$. Similarly, the computation of $\mathrm{SQM}_{\epsilon,\gamma}$ 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: $0, 1, 0, 1, 0, 1, \ldots$. Let the sequence be denoted by $a_{\mathrm{SQM}_{\epsilon,\gamma}}(j)$, the formula of the sequence is $a_{\mathrm{SQM}_{\epsilon,\gamma}}(j) = j \bmod 2$. If pairing all blocks in $\mathrm{BM}_{\nu=2,\epsilon,\gamma}$ and all quantiles in $\mathrm{SQM}_{\epsilon,\gamma}$, there are two possible pairings of $a_{\mathrm{BM}_{\nu=2}}(j)$ and $a_{\mathrm{SQM}_{\epsilon,\gamma}}(j)$. One pairing occurs when $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j) = a_{\mathrm{SQM}_{\epsilon,\gamma}}(j) = 1$, while the other involves the sequence $0, 1, 0$ from $a_{\mathrm{BM}_{\nu=2,\epsilon,\gamma}}(j)$ paired with $1, 0, 1$ from $a_{\mathrm{SQM}_{\epsilon,\gamma}}(j)$. By leveraging the same principle as Theorem .13 and the second γ-orderliness (replacing the two quantile averages with one quantile average between them), the desired result follows. □

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

Hodges–Lehmann inequality and γ-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 (8) 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, (8) and was later independently discovered by Sen (1963) (36, 37). 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 (38). Serfling further advanced the understanding by generalizing the H-L kernel as $hl_k(x_1, \ldots, x_k) = \frac{1}{k}\sum_{i=1}^{k} x_i$, where $k \in \mathbb{N}$ (38). Here, the weighted H-L kernel is defined as $whl_k(x_1, \ldots, x_k) = \frac{\sum_{i=1}^{k} x_i \mathbf{w}_i}{\sum_{i=1}^{k} \mathbf{w}_i}$, where $\mathbf{w}_i$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:

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

where $L_{\epsilon_0,\gamma,n}(Y)$ represents the ϵ_0,γ-L-estimator that uses the sorted sequence, $\mathrm{sort}\left(\left(whl_k\left(X_{N_1}, \cdots, X_{N_k}\right)\right)_{N=1}^{\binom{n}{k}}\right)$, as input. The upper asymptotic breakdown point of $LL_{k,\epsilon,\gamma}$ is $\epsilon = 1 - (1-\epsilon_0)^{\frac{1}{k}}$, as proven in DSSM II. There are two ways to adjust the breakdown point: either by setting k as a constant and adjusting ϵ_0, or by setting ϵ_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 $(1 - \lceil k \rceil + k)b$ times with each sample size of $\lceil k \rceil$, $(1 - k + \lfloor k \rfloor)b \in \mathbb{N}$, $(1 - \lceil k \rceil + k)b \in \mathbb{N}$. 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, $whl_{k=n}(X_1, \ldots, X_n)$. When $\mathbf{w}_i = 1$, 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 $\mathbf{w}_i$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 $k \to n$. Therefore, if $\frac{\sum_{i=1}^{n} X_i \mathbf{w}_i}{\sum_{i=1}^{n} \mathbf{w}_i}$ approaches the population mean when $n \to \infty$, all LL-statistics based on such consistent kernel function approach the population mean as $k \to \infty$. For example, if $whl_k = \mathrm{BM}_{\nu,\epsilon_k,n=k}$, $\nu \ll \epsilon_k^{-1}$, $\epsilon_k \to 0$, such kernel function is consistent. These cases are termed the LL-mean ($\mathrm{LLM}_{k,\epsilon,\gamma,n}$). By substituting the $\mathrm{WA}_{\epsilon_0,\gamma,n}$ for the $L_{\epsilon_0,\gamma,n}$ in LL-statistic, the resulting statistic is referred to as the weighted L-statistic ($\mathrm{WL}_{k,\epsilon,\gamma,n}$). The case having a consistent kernel function is termed as the weighted L-mean ($\mathrm{WLM}_{k,\epsilon,\gamma,n}$). The $w_i = 1$ case of $\mathrm{WLM}_{k,\epsilon,\gamma,n}$ is termed the weighted Hodges-Lehmann mean ($\mathrm{WHLM}_{k,\epsilon,\gamma,n}$). The $\mathrm{WHLM}_{k=1,\epsilon,\gamma,n}$ is the weighted average. If $k \geq 2$ and the WA in WHLM is set as TM_{ϵ_0}, it is called the trimmed H-L mean (Figure **??**, $k = 2$, $\epsilon_0 = \frac{15}{64}$). The $\mathrm{THLM}_{k=2,\epsilon,\gamma=1,n}$ appears similar to the Wilcoxon's one-sample statistic investigated by Saleh in 1976 (39), 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 $\mathrm{THLM}_{k=2,\epsilon=1-\left(1-\frac{1}{2}\right)^{\frac{1}{2}},\gamma=1,n}$ is the Hodges-Lehmann estimator, or more generally, a special case of the median Hodges-Lehmann mean ($m\mathrm{HLM}_{k,n}$). $m\mathrm{HLM}_{k,n}$ is asymptotically equivalent to the $\mathrm{MoM}_{k,b=\frac{n}{k}}$ as discussed previously, Therefore, it is possible to define a series of location estimators, analogous to the WHLM, based on MoM. For example, the γ-median of means, $\gamma m\mathrm{oM}_{k,b=\frac{n}{k},n}$, is defined by replacing the median in $\mathrm{MoM}_{k,b=\frac{n}{k},n}$ with the γ-median.

The hl_k kernel distribution, denoted as F_{hl_k}, can be defined as the probability distribution of the sorted sequence $\mathrm{sort}\left(\left(hl_k\left(X_{N_1}, \cdots, X_{N_k}\right)\right)_{N=1}^{\binom{n}{k}}\right)$. For any real value y, the cdf of the hl_k kernel distribution is given by: $F_{h_{\mathbf{k}}}(y) = \Pr(Y_i \leq y)$, 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 γ-U-orderliness:

$$(\forall k_2 \geq k_1 \geq 1, \gamma m\mathrm{HLM}_{k_2, \epsilon=1-\left(\frac{\gamma}{1+\gamma}\right)^{\frac{1}{k_2}}, \gamma} \geq \gamma m\mathrm{HLM}_{k_1, \epsilon=1-\left(\frac{\gamma}{1+\gamma}\right)^{\frac{1}{k_1}}, \gamma}) \vee$$

$$(\forall k_2 \geq k_1 \geq 1, \gamma m\mathrm{HLM}_{k_2, \epsilon=1-\left(\frac{\gamma}{1+\gamma}\right)^{\frac{1}{k_2}}, \gamma} \leq \gamma m\mathrm{HLM}_{k_1, \epsilon=1-\left(\frac{\gamma}{1+\gamma}\right)^{\frac{1}{k_1}}, \gamma})$$

where $\gamma m\mathrm{HLM}_k$ sets the WA in WHLM as γ-median, with γ being constant. The direction of the inequality depends on the relative magnitudes of $\gamma m\mathrm{HLM}_{k=1,\epsilon,\gamma} = \gamma m$ and $\gamma m\mathrm{HLM}_{k=\infty,\epsilon,\gamma} = \mu$. The Hodges-Lehmann inequality can be defined as a special case of the γ-U-orderliness when $\gamma = 1$. When $\gamma \in \{0, \infty\}$, the γ-U-orderliness is valid for any distribution as previously shown. If $\gamma \notin \{0, \infty\}$, analytically proving the validity of the γ-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 $\int_0^{\text{H-L}} (f_{hl_2}(s))\, ds = \frac{1}{2}$. For the exponential distribution, $f_{hl_2,exp}(x) = 4\lambda^{-2} x e^{-2\lambda^{-1}x}$, λ is a scale parameter, $E[\text{H-L}] = \frac{-W_{-1}\left(-\frac{1}{2e}\right)-1}{2}\lambda \approx 0.839\lambda$, 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 γ-U-orderliness is bounded under certain assumptions, as shown below.

Theorem .15. *For any distribution with a finite second central moment, σ^2, the following concentration bound can be established for the γ-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 $\widehat{\mu}_i$, $1 \leq i \leq b$. Observe that the event $\left\{\gamma mo\mathrm{M}_{k,b=\frac{n}{k},n} - \mu > \frac{t\sigma}{\sqrt{k}}\right\}$ necessitates the condition that there are at least $b(1 - \frac{\gamma}{1+\gamma})$ of $\widehat{\mu}_i$s larger than μ by more than $\frac{t\sigma}{\sqrt{k}}$, i.e., $\left\{\gamma mo\mathrm{M}_{k,b=\frac{n}{k},n} - \mu > \frac{t\sigma}{\sqrt{k}}\right\} \subset \left\{\sum_{i=1}^{b} \mathbf{1}_{\left(\widehat{\mu}_i - \mu\right) > \frac{t\sigma}{\sqrt{k}}} \geq b\left(1 - \frac{\gamma}{1+\gamma}\right)\right\}$, where $\mathbf{1}_A$ is the indicator of event A. Assuming a finite second central moment, σ^2, it follows from one-sided Chebeshev's inequality that $\mathbb{E}\left(\mathbf{1}_{\left(\widehat{\mu}_i - \mu\right) > \frac{t\sigma}{\sqrt{k}}}\right) = \mathbb{P}\left(\left(\widehat{\mu}_i - \mu\right) > \frac{t\sigma}{\sqrt{k}}\right) \leq \frac{\sigma^2}{k\sigma^2 + t^2\sigma^2}$.

Given that $\mathbf{1}_{\left(\widehat{\mu}_i - \mu\right) > \frac{t\sigma}{\sqrt{k}}} \in [0,1]$ 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, $\mathbb{P}\left(\gamma mo\mathrm{M}_{k,b=\frac{n}{k},n} - \mu > \frac{t\sigma}{\sqrt{k}}\right) \leq \mathbb{P}\left(\sum_{i=1}^{b} \mathbf{1}_{\left(\widehat{\mu}_i - \mu\right) > \frac{t\sigma}{\sqrt{k}}} \geq b\left(1 - \frac{\gamma}{1+\gamma}\right)\right) = \mathbb{P}\left(\frac{1}{b}\sum_{i=1}^{b}\left(\mathbf{1}_{\left(\widehat{\mu}_i - \mu\right) > \frac{t\sigma}{\sqrt{k}}} - \mathbb{E}\left(\mathbf{1}_{\left(\widehat{\mu}_i - \mu\right) > \frac{t\sigma}{\sqrt{k}}}\right)\right) \geq \left(1 - \frac{\gamma}{1+\gamma}\right) - \mathbb{E}\left(\mathbf{1}_{\left(\widehat{\mu}_i - \mu\right) > \frac{t\sigma}{\sqrt{k}}}\right)\right) \leq e^{-2b\left(\left(1 - \frac{\gamma}{1+\gamma}\right) - \mathbb{E}\left(\mathbf{1}_{\left(\widehat{\mu}_i - \mu\right) > \frac{t\sigma}{\sqrt{k}}}\right)\right)^2} \leq e^{-2b\left(1 - \frac{\gamma}{1+\gamma} - \frac{\sigma^2}{k\sigma^2 + t^2\sigma^2}\right)^2} = e^{-2b\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)^2}$. □

Theorem .16. *Let $B(k,\gamma,t,n) = e^{-\frac{2n}{k}\left(\frac{1}{1+\gamma} - \frac{1}{k+t^2}\right)^2}$. If $n \in \mathbb{N}$, $\gamma \geq 0$, $0 \leq t^2 < \gamma + 1$, and $\gamma - t^2 + 1 \leq k \leq \frac{1}{2}\sqrt{9\gamma^2 + 18\gamma - 8\gamma t^2 - 8t^2 + 9} + \frac{1}{2}\left(3\gamma - 2t^2 + 3\right)$, B is monotonic decreasing with respect to k.*

Proof. Since $\frac{\partial B}{\partial k} = \left(\frac{2n\left(\frac{1}{\gamma+1} - \frac{1}{k+t^2}\right)^2}{k^2} - \frac{4n\left(\frac{1}{\gamma+1} - \frac{1}{k+t^2}\right)}{k\left(k+t^2\right)^2}\right) e^{-\frac{2n\left(\frac{1}{\gamma+1} - \frac{1}{k+t^2}\right)^2}{k}}$ and $n \in \mathbb{N}$, $\frac{\partial B}{\partial k} \leq 0 \Leftrightarrow \frac{2n\left(\frac{1}{\gamma+1} - \frac{1}{k+t^2}\right)^2}{k^2} - \frac{4n\left(\frac{1}{\gamma+1} - \frac{1}{k+t^2}\right)}{k\left(k+t^2\right)^2} \leq 0 \Leftrightarrow \frac{2n\left(-\gamma+k+t^2-1\right)\left(k^2 - 3(\gamma+1)k + 2kt^2 + t^2\left(-\gamma+t^2-1\right)\right)}{(\gamma+1)^2 k^2 \left(k+t^2\right)^3} \leq 0 \Leftrightarrow \left(-\gamma + k + t^2 - 1\right)\left(k^2 - 3(\gamma+1)k + 2kt^2 + t^2\left(-\gamma + t^2 - 1\right)\right) \leq 0$. When the factors are expanded, it yields a cubic inequality in terms of k: $k^3 + k^2\left(3t^2 - 4(\gamma+1)\right) + 3k\left(\gamma - t^2 + 1\right)^2 + t^2\left(\gamma - t^2 + 1\right)^2 \leq 0$. 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 $\gamma - t^2 + 1 \leq k \leq \frac{1}{2}\sqrt{9\gamma^2 + 18\gamma - 8\gamma t^2 - 8t^2 + 9} + \frac{1}{2}\left(3\gamma - 2t^2 + 3\right)$. □

Let X be a random variable and $\bar{Y} = \frac{1}{k}(Y_1 + \cdots + Y_k)$ be the average of k independent, identically distributed copies of X. Applying the variance operation gives: $\mathrm{Var}(\bar{Y}) = \mathrm{Var}\left(\frac{1}{k}(Y_1 + \cdots + Y_k)\right) = \frac{1}{k^2}(\mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_k)) = \frac{1}{k^2}(k\sigma^2) = \frac{\sigma^2}{k}$, 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., $\mathrm{Var}(cX) = E[(cX - E[cX])^2] = E[(cX - cE[X])^2] = E[c^2(X - E[X])^2] = c^2 E[((X) - E[X])^2] = c^2\mathrm{Var}(X)$. 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, σ^2,(34), a conservative asymptotic bias bound of $\gamma mo\mathrm{M}_{k,b=\frac{n}{k}}$ can be established as $\gamma mo\mathrm{M}_{k,b=\frac{n}{k}} - \mu \leq \sqrt{\frac{\frac{\gamma}{1+\gamma}}{1 - \frac{\gamma}{1+\gamma}}}\sigma_{hl_k} = \sqrt{\frac{\gamma}{k}}\sigma$. That implies in Theorem .15, $t < \sqrt{\gamma}$, so when $\gamma = 1$, the upper bound of k, subject to the monotonic decreasing constraint, is $2 + \sqrt{5} < \frac{1}{2}\sqrt{9 + 18 - 8t^2 - 8t^2 + 9} + \frac{1}{2}\left(3 - 2t^2 + 3\right) \leq 6$, the lower bound is $1 < 2 - t^2 \leq 2$. These analyses elucidate a surprising result: although the conservative asymptotic bound of $\mathrm{MoM}_{k,b=\frac{n}{k}}$ 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 νth γ-ordered if and only if the original distribution is νth γ-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 (40, 41). Theorem .9 implies that the violation of νth γ-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 νth γ-ordered and the distribution itself is νth γ-ordered and γ-U-ordered, then the distribution is called νth γ-U-ordered. The following theorems highlight the significance of γ-symmetric distributions.

Theorem .17. *Any γ-symmetric distribution is νth γ-U-ordered, provided that the γ is the same.*

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 .18. $whl_k\left(x_1 = \lambda x_1 + \mu, \ldots, x_k = \lambda x_k + \mu\right) = \lambda whl_k\left(x_1, \ldots, x_k\right) + \mu$.

Proof. $whl_k\left(x_1 = \lambda x_1 + \mu, \cdots, x_k = \lambda x_k + \mu\right) = \frac{\sum_{i=1}^{k}(\lambda x_i + \mu) w_i}{\sum_{i=1}^{k} w_i} = \frac{\sum_{i=1}^{k}\lambda x_i w_i + \sum_{i=1}^{k}\mu w_i}{\sum_{i=1}^{k} w_i} = \lambda\frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i} + \frac{\sum_{i=1}^{k}\mu w_i}{\sum_{i=1}^{k} w_i} = \lambda\frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i} + \mu = \lambda whl_k\left(x_1, \cdots, x_k\right) + \mu$. □

According to Theorem .18, the γ-weighted inequality for a right-skewed distribution can be modified as $\forall 0 \leq \epsilon_{0_1} \leq \epsilon_{0_2} \leq \frac{1}{1+\gamma}$, $\mathrm{WLM}_{k,\epsilon=1-\left(1-\epsilon_{0_1}\right)^{\frac{1}{k}},\gamma} \geq \mathrm{WLM}_{k,\epsilon=1-\left(1-\epsilon_{0_2}\right)^{\frac{1}{k}},\gamma}$, which holds the same rationale as the γ-weighted inequality defined in the last section. If the νth γ-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 **??**), while its maximum breakdown point is ≈ 0.065 if $\nu = 3$. A comparison of the biases of $\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{8}}$, $\mathrm{SQM}_{\epsilon=\frac{1}{8}}$, $\mathrm{THLM}_{k=2,\epsilon=\frac{1}{8}}$, $\mathrm{WHLM}_{k=2,\epsilon=\frac{1}{8}}$, $\mathrm{MHHLM}_{k=\frac{2\ln(2)-\ln(3)}{3\ln(2)-\ln(7)},\epsilon=\frac{1}{8}}$ (midhinge H-L mean), $m\mathrm{HLM}_{k=\frac{\ln(2)}{3\ln(2)-\ln(7)},\epsilon=\frac{1}{8}}$, $\mathrm{THLM}_{k=5,\epsilon=\frac{1}{8}}$, and $\mathrm{WHLM}_{k=5,\epsilon=\frac{1}{8}}$ is appropriate (Figure **??**, SI Dataset S1), given their same breakdown points, with $m\mathrm{HLM}_{k=\frac{\ln(2)}{3\ln(2)-\ln(7)},\epsilon=\frac{1}{8}}$ 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)'s seminal work that MoM is nearly optimal with regards to concentration bounds for heavy-tailed distributions (15).

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 (42). Among various low-discrepancy sequences, Sobol sequences are often favored in quasi-Monte Carlo methods (43). 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 (44). By using a deterministic approach, the variance of $m\mathrm{HLM}_{k,n}$ is much lower than that of $\mathrm{MoM}_{k,b=\frac{n}{k}}$ (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.

1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae.* (Henricus Dieterich), (1823).
2. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
3. P Daniell, Observations weighted according to order. *Am. J. Math.* **42**, 222–236 (1920).
4. JW Tukey, A survey of sampling from contaminated distributions in *Contributions to probability and statistics.* (Stanford University Press), pp. 448–485 (1960).
5. WJ Dixon, Simplified Estimation from Censored Normal Samples. *The Annals Math. Stat.* **31**, 385 – 391 (1960).
6. K Danielak, T Rychlik, Theory & methods: Exact bounds for the bias of trimmed means. *Aust. & New Zealand J. Stat.* **45**, 83–96 (2003).
7. M Bieniek, Comparison of the bias of trimmed and winsorized means. *Commun. Stat. Methods* **45**, 6641–6650 (2016).
8. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
9. F Wilcoxon, Individual comparisons by ranking methods. *Biom. Bull.* **1**, 80–83 (1945).
10. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
11. Q Sun, WX Zhou, J Fan, Adaptive huber regression. *J. Am. Stat. Assoc.* **115**, 254–265 (2020).
12. T Mathieu, Concentration study of m-estimators using the influence function. *Electron. J. Stat.* **16**, 3695–3750 (2022).
13. AS Nemirovskij, DB Yudin, *Problem complexity and method efficiency in optimization.* (Wiley-Interscience), (1983).
14. D Hsu, S Sabato, Heavy-tailed regression with a generalized median-of-means in *International Conference on Machine Learning.* (PMLR), pp. 37–45 (2014).
15. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.* **44**, 2695–2725 (2016).
16. P Laforgue, S Clémençon, P Bertail, On medians of (randomized) pairwise means in *International Conference on Machine Learning.* (PMLR), pp. 1272–1281 (2019).
17. G LECUÉ, M LERASLE, Robust machine learning by median-of-means: Theory and practice. *The Annals Stat.* **48**, 906–931 (2020).
18. B Efron, Bootstrap methods: Another look at the jackknife. *The Annals Stat.* **7**, 1–26 (1979).
19. PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. *The annals statistics* **9**, 1196–1217 (1981).
20. PJ Bickel, DA Freedman, Asymptotic normality and the bootstrap in stratified sampling. *The annals statistics* **12**, 470–482 (1984).
21. R Helmers, P Janssen, N Veraverbeke, *Bootstrapping U-quantiles.* (CWI. Department of Operations Research, Statistics, and System Theory [BS]), (1990).
22. 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).
23. G McIntyre, A method for unbiased selective sampling, using ranked sets. *Aust. journal agricultural research* **3**, 385–390 (1952).
24. DL Donoho, PJ Huber, The notion of breakdown point. *A festschrift for Erich L. Lehmann* **157184** (1983).
25. 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).
26. PJ Bickel, On adaptive estimation. *The Annals Stat.* **10**, 647–671 (1982).
27. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparametric models.* (Springer) Vol. 4, (1993).
28. JT Runnenburg, Mean, median, mode. *Stat. Neerlandica* **32**, 73–79 (1978).
29. Wv Zwet, Mean, median, mode ii. *Stat. Neerlandica* **33**, 1–5 (1979).
30. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
31. 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).
32. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
33. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).
34. 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).
35. JW Tukey, *Exploratory data analysis.* (Reading, MA) Vol. 2, (1977).
36. PK Sen, On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. *Biometrics* pp. 532–552 (1963).
37. M Ghosh, MJ Schell, PK Sen, A conversation with pranab kumar sen. *Stat. Sci.* pp. 548–564 (2008).
38. RJ Serfling, Generalized l-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).
39. A Ehsanes Saleh, Hodges-lehmann estimate of the location parameter in censored samples. *Annals Inst. Stat. Math.* **28**, 235–247 (1976).
40. J Hodges, E Lehmann, Matching in paired comparisons. *The Annals Math. Stat.* **25**, 787–791 (1954).
41. K Chung, Sur les lois de probabilité unimodales. *COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES* **236**, 583–584 (1953).

42. 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).
43. 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).
44. KA Do, P Hall, Quasi-random resampling for the bootstrap. *Stat. Comput.* **1**, 13–22 (1991).