Near-consistent robust estimations of moments for unimodal distributions

Tuban Lee[a,1]

[a]Institute of Biomathematics, Macau SAR 999078, China

This manuscript was compiled on March 18, 2023

Descriptive statistics for parametric models currently heavily rely on the accuracy of distributional assumptions. Here, based on the invariant structures of unimodal distributions, a series of sophisticated yet efficient estimators, robust to both gross errors and departures from parametric assumptions, are proposed for estimating mean and central moments with insignificant asymptotic biases for common continuous unimodal distributions. This article also illuminates the understanding of the common nature of probability distributions and the measures of them.

orderliness | invariant | unimodal | adaptive estimation | U-statistics

The asymptotic inconsistencies between sample mean ($\bar{x}$) and nonparametric robust location estimators in asymmetric distributions on the real line have been noticed for more than two centuries (1), yet remain unsolved. Strictly speaking, it is unsolvable as by trimming, some information about the original distribution is removed, making it impossible to estimate the values of the removed parts without distributional assumptions. Newcomb (1886, 1912) provided the first modern approach to this problem by developing a class of estimators that gives "less weight to the more discordant observations" (2, 3). In 1964, Huber (4) used the minimax procedure to obtain M-estimator for the contaminated normal distribution, which has played a pre-eminent role in the later development of robust statistics. However, as previously demonstrated, under growing asymmetric departures from normality, the bias of the Huber M-estimator increases rapidly. This is a common issue in parametric estimations. For example, He and Fung (1999) constructed (5) a robust M-estimator for the two-parameter Weibull distribution, from which all moments can be calculated. Nonetheless, it is inadequate for the gamma, Perato, lognormal, and the generalized Gaussian distributions (SI Dataset S1). Another old and interesting approach is arithmetically computing the parameters using one or more L-statistics as inputs, such as percentile estimators. Examples of percentile estimators for the Weibull distribution, the reader is referred to Menon (1963) (6), Dubey (1967) (7), Hassanein (1971) (8), Marks (2005) (9), and Boudt, Caliskan, and Croux (2011) (10)'s works. At the outset of the study of percentile estimators, it was known that they arithmetically utilizes the invariant structures of probability distributions (6, 11, 12). Maybe such estimators can be named as I-statistics. Formally, an estimator is classified as an I-statistic if it asymptotically satisfies $\mathrm{I}\left(\mathrm{LE}_1, \cdots, \mathrm{LE}_l\right) = \left(\theta_1, \cdots, \theta_q\right)$ for the distribution it is consistent, where LEs are calculated with the use of L-statistics, I is defined using arithmetic operations and constants, but it may also incorporate other functions, and θs are the population parameters it estimates. A subclass of I-statistics, arithmetic I-statistics, is defined as LEs are L-statistics, I is solely defined using arithmetic operations and constants. Since some percentile estimators use the logarithmic function to transform all random variables before computing the L-statistics, a percentile estimator might not always be an arithmetic I-statistic (7). In this article, two subclasses of I-statistics are introduced, arithmetic I-statistics and quantile I-statistics. Examples of quantile I-statistics will be discussed later. Based on L-statistics, I-statistics are naturally robust. Compared to probability density functions (pdfs) and cumulative distribution functions (cdfs), the quantile functions of many parametric distributions are more elegant. Since the expectation of an L-statistic can be expressed as an integral of the quantile function, I-statistics are often analytically obtainable. However, the performance of the aforementioned examples is often worse than that of the robust M-statistics when the distributional assumption is violated (SI Dataset S1). Even when distributions such as the Weibull and gamma belong to the same larger family, the generalized gamma distribution, a misassumption can still result in substantial biases, rendering the approach ill-suited.

In previous research on semiparametric robust mean estimation, the binomial mean (BM_ϵ) is still inconsistent for any skewed distribution, despite having much smaller asymptotic biases than other weighted averages. All robust location estimators commonly used are symmetric due to the universality of the symmetric distributions. One can construct an asymmetric weighted average that is consistent for a semiparametric class of skewed distributions. This approach has been investigated previously, but its lack of symmetry makes it suitable only for certain applications (13). Shifting from semiparametrics to parametrics, an ideal robust location estimator would have a non-sample-dependent breakdown point (defined in Subsection **??**) and be consistent for any symmetric distribution and a skewed distribution with finite second moments. This is called an invariant mean. Based on the mean-symmetric weighted

Significance Statement

Bias, variance, and contamination are the three main errors in statistics. Consistent robust estimation is unattainable without parametric assumptions. Here, based on a paradigm shift inspired by mean-median-mode inequality, Bickel-Lehmann spread, and adaptive estimation, invariant moments are proposed as a means of achieving near-consistent and robust estimations of moments, even in scenarios where moderate violations of distributional assumptions occur, while the variances are sometimes smaller than those of the sample moments.

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

average-median inequality, the recombined mean is defined as

$$rm_{d,\epsilon,n} := \lim_{c\to\infty}\left(\frac{(\mathrm{SWA}_{\epsilon,n}+c)^{d+1}}{(median+c)^{d}}-c\right),$$

where d is the key factor for bias correction, $\mathrm{SWA}_{\epsilon,n}$ is $\mathrm{BM}_{\epsilon,n}$ in the first three Subsections, but other symmetric weighted averages can also be used in practice as long as the inequalities hold. The following theorem shows the significance of this arithmetic I-statistic.

Theorem .1. *If the second moments are finite, $rm_{d\approx 0.375,\epsilon=\frac{1}{8}}$ is a consistent mean estimator for the exponential and any symmetric distributions and the Pareto distribution with quantile function $Q(p)=x_m(1-p)^{-\frac{1}{\alpha}}$, $x_m>0$, when $\alpha\to\infty$.*

Proof. Finding d and ϵ that make $rm_{d,\epsilon}$ a consistent mean estimator is equivalent to finding the solution of $E[rm_{d,\epsilon}]=E[X]$. Rearranging the definition, $rm_{d,\epsilon}=\lim_{c\to\infty}\left(\frac{(\mathrm{BM}_\epsilon+c)^{d+1}}{(median+c)^{d}}-c\right)=(d+1)\,\mathrm{BM}_\epsilon-d\mathrm{median}=\mu$. So, $d=\frac{\mu-\mathrm{BM}_\epsilon}{\mathrm{BM}_\epsilon-median}$. The quantile function of the exponential distribution is $Q(p)=\ln\left(\frac{1}{1-p}\right)\lambda$. $E[x]=\lambda$. $E[median]=Q\left(\frac{1}{2}\right)=\ln 2\lambda$. For the exponential distribution, the expectation of $\mathrm{BM}_{\frac{1}{8}}$ is $E\left[\mathrm{BM}_{\frac{1}{8}}\right]=\lambda\left(1+\ln\left(\frac{46656}{8575\sqrt{35}}\right)\right)$. Obviously, the scale parameter λ can be canceled out, $d\approx 0.375$. The proof of the second assertion follows directly from the coincidence property. For any symmetric distribution with a finite second moment, $E[\mathrm{BM}_\epsilon]=E[median]=E[X]$. Then $E[rm_{d,\epsilon}]=\lim_{c\to\infty}\left(\frac{(E[X]+c)^{d+1}}{(E[X]+c)^{d}}-c\right)=E[X]$. The proof for the Pareto distribution is more general. The mean of the Pareto distribution is given by $\frac{\alpha x_m}{\alpha-1}$. The d value with two unknown percentiles p_1 and p_2 for the Pareto distribution is $d_{Perato}=\frac{\frac{\alpha x_m}{\alpha-1}-x_m(1-p_1)^{-\frac{1}{\alpha}}}{x_m(1-p_1)^{-\frac{1}{\alpha}}-x_m(1-p_2)^{-\frac{1}{\alpha}}}$. Since any weighted average can be expressed as an integral of the quantile function, $\lim_{\alpha\to\infty}\frac{\frac{\alpha}{\alpha-1}-(1-p_1)^{-1/\alpha}}{(1-p_1)^{-1/\alpha}-(1-p_2)^{-1/\alpha}}=-\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$, the d value for the Pareto distribution approaches that of the exponential distribution as $\alpha\to\infty$, regardless of the type of weighted average used. This completes the demonstration. $\square$

Theorem .1 implies that for the Weibull, gamma, Pareto, lognormal and generalized Gaussian distribution, $rm_{d\approx 0.375,\epsilon=\frac{1}{8}}$ is consistent for at least one particular case of these two-parameter distributions. The biases of $rm_{d\approx 0.375,\epsilon=\frac{1}{8}}$ for distributions with skewness between those of the exponential and symmetric distributions are tiny (SI Dataset S1). $rm_{d\approx 0.375,\epsilon=\frac{1}{8}}$ exhibits excellent performance for all these common unimodal distributions (SI Dataset S1).

Besides introducing the concept of invariant mean, the purpose of this paper is to demonstrate that, in light of previous works, the estimation of central moments can be transformed into a location estimation problem by using U-statistics, the central moment kernel distributions possess desirable properties, and a series of sophisticated yet efficient robust estimators can be constructed whose biases are typically smaller than the variances (as seen in Table **??** for $n=5400$) for unimodal distributions.

Background and Main Results

A. Invariant mean. It has long been known that a theoretical model can be adjusted to fit the first two moments of the observed data. A continuous distribution belonging to a location–scale family takes the form $F(x)=F_0\left(\frac{x-\mu}{\lambda}\right)$, where F_0 is a "standard" distribution. Therefore, $F(x)=Q^{-1}(x)\to x=Q(p)=\lambda Q_0(p)+\mu$. Thus, any weighted average can be expressed as $\lambda\mathrm{WA}_0(\epsilon)+\mu$, where $\mathrm{WA}_0(\epsilon)$ is an integral of $Q_0(p)$ according to the definition of the weighted average. The simultaneous cancellation of μ and λ in $\frac{(\lambda\mu_0+\mu)-(\lambda\mathrm{BM}_0(\epsilon)+\mu)}{(\lambda\mathrm{BM}_0(\epsilon))+\mu-(\lambda median_0+\mu)}$ assures that d is a constant. Consequently, the roles of BM_ϵ and median in $rm_{d,\epsilon}$ can be replaced by any weighted averages, although only symmetric weighted averages are considered in defining the invariant mean.

The performance in heavy-tailed distributions can be further improved by constructing the quantile mean as

$$qm_{d,\epsilon,n} := \hat{Q}_n\left(\left(\hat{F}_n\left(\mathrm{SWA}_{\epsilon,n}\right)-\frac{1}{2}\right)d+\hat{F}_n\left(\mathrm{SWA}_{\epsilon,n}\right)\right),$$

provided that $\hat{F}_n\left(\mathrm{SWA}_{\epsilon,n}\right)\geq\frac{1}{2}$, where $\hat{F}_n(x)$ is the empirical cumulative distribution function of the sample, $\hat{Q}_n$ is the sample quantile function. The most popular method for computing the sample quantile function was proposed by Hyndman and Fan in 1996 (14). To minimize the finite sample bias, here, $\hat{F}_n(x):=\frac{1}{n}\left(\frac{x-X_{sp}}{X_{sp+1}-X_{sp}}+sp\right)$, where $sp=\sum_{i=1}^{n}\mathbf{1}_{X_i\leq x}$, $\mathbf{1}_A$ is the indicator of event A. The solution of $\hat{F}_n\left(\mathrm{SWA}_{\epsilon,n}\right)<\frac{1}{2}$ is reversing the percentile by $1-\hat{F}_n\left(\mathrm{SWA}_{\epsilon,n}\right)$, the obtained percentile is also reversed. Without loss of generality, in the following discussion, only the case where $\hat{F}_n\left(\mathrm{SWA}_{\epsilon,n}\right)\geq\frac{1}{2}$ is considered. Moreover, in extreme heavy-tailed distributions, the calculated percentile can exceed the breakdown point of SWA_ϵ, so the percentile will be modified to $1-\epsilon$ if this occurs. The quantile mean uses the location-scale invariant in a different way as shown in the following proof.

Theorem A.1. *$qm_{d\approx 0.321,\epsilon=\frac{1}{8}}$ is a consistent mean estimator for the exponential, Pareto ($\alpha\to\infty$) and any symmetric distributions provided that the second moments are finite.*

Proof. Similarly, rearranging the definition, $d=\frac{F(\mu)-F(\mathrm{BM}_\epsilon)}{F(\mathrm{BM}_\epsilon)-\frac{1}{2}}$. The cdf of the exponential distribution is $F(x)=1-e^{-\lambda^{-1}x}$, $\lambda\geq 0$, $x\geq 0$, the expectation of BM_ϵ can be expressed as $\lambda\mathrm{BM}_0(\epsilon)$, so $F(\mathrm{BM}_\epsilon)$ is free of λ. When $\epsilon=\frac{1}{8}$, $d=\frac{-e^{-1}+e^{-\left(1+\ln\left(\frac{46656}{8575\sqrt{35}}\right)\right)}}{\frac{1}{2}-e^{-\left(1+\ln\left(\frac{46656}{8575\sqrt{35}}\right)\right)}}\approx 0.321$. The proof of the symmetric case is similar. Since for any symmetric distribution with a finite second moment, $F(E[\mathrm{BM}_\epsilon])=F(\mu)=\frac{1}{2}$. Then, the expectation of the quantile mean is $qm_{d,\epsilon}=F^{-1}\left(\left(F(\mu)-\frac{1}{2}\right)d+F(\mu)\right)=F^{-1}(0+F(\mu))=\mu$.

For the assertion related to the Pareto distribution, the cdf of it is $1-\left(\frac{x_m}{x}\right)^{\alpha}$. So, the d value with two unknown percentile p_1 and p_2 is

$$d_{Pareto}=\frac{1-\left(\frac{x_m}{\frac{\alpha x_m}{\alpha-1}}\right)^{\alpha}-\left(1-\left(\frac{x_m}{x_m(1-p_1)^{-\frac{1}{\alpha}}}\right)^{\alpha}\right)}{\left(1-\left(\frac{x_m}{x_m(1-p_1)^{-\frac{1}{\alpha}}}\right)^{\alpha}\right)-\left(1-\left(\frac{x_m}{x_m(1-p_2)^{-\frac{1}{\alpha}}}\right)^{\alpha}\right)}=$$

$\frac{1-\left(\frac{\alpha-1}{\alpha}\right)^{\alpha}-p_1}{p_1-p_2}$. When $\alpha\to\infty$, $\left(\frac{\alpha-1}{\alpha}\right)^{\alpha}=\frac{1}{e}$. The d value for the exponential distribution is identical, since $d_{exp}=$

$\frac{\left(1-e^{-1}\right)-\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)}{\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)-\left(1-e^{-\ln\left(\frac{1}{1-p_2}\right)}\right)} = \frac{1-\frac{1}{e}-p_1}{p_1-p_2}$. All results are now proven. □

The definitions of location and scale parameters are such that they must satisfy $F(x;\lambda,\mu) = F(\frac{x-\mu}{\lambda};1,0)$. By recalling $x = \lambda Q_0(p) + \mu$, it follows that the percentile of any weighted average is free of λ and μ, which guarantees the validity of the quantile mean. The quantile mean is a quantile I-statistic. Specifically, an estimator is classified as a quantile I-statistic if LEs are percentiles of a distribution obtained by plugging L-statistics into a cumulative distribution function and I is defined with arithmetic operations, constants and quantile functions. $qm_{d\approx0.321,\epsilon=\frac{1}{8}}$ works better in the fat-tail scenarios (SI Dataset S1). Theorem .1 and A.1 show that $rm_{d\approx0.375,\epsilon=\frac{1}{8}}$ and $qm_{d\approx0.321,\epsilon=\frac{1}{8}}$ are both consistent mean estimators for any symmetric distribution and a skewed distribution with finite second moments. It's obvious that the breakdown points of $rm_{d\approx0.375,\epsilon=\frac{1}{8}}$ and $qm_{d\approx0.321,\epsilon=\frac{1}{8}}$ are both $\frac{1}{8}$. Therefore they are all invariant means.

To study the impact of the choice of SWAs in rm and qm, it is constructive to recall that a symmetric weighted average is a linear combination of symmetric quantile averages. While using a less-biased symmetric weighted average can generally enhance performance (SI Dataset S1), there is a greater risk of violation in the semiparametric framework. However, the mean-SWA-median inequality is robust to slight fluctuations of the SQA function of the underlying distribution. Suppose the SQA function is generally decreasing in $[0,u]$, but increasing in $[u,\frac{1}{2}]$, since $1-2\epsilon$ of the symmetric quantile averages will be included in the computation of SWA_ϵ, as long as $\frac{1}{2}-u \ll 1-2\epsilon$, and other portions of the SQA function satisfy the inequality constraints that define the νth orderliness on which the SWA_ϵ is based, the mean-SWA_ϵ-median inequality will still hold. This is due to the violation being bounded (15) and therefore cannot be extreme for unimodal distributions. For instance, the SQA function is non-monotonic when the shape parameter of the Weibull distribution $\alpha > \frac{1}{1-\ln(2)} \approx 3.259$ as shown in the previous article, the violation of the third orderliness starts near this parameter as well, yet the mean-$\mathrm{BM}_{\frac{1}{8}}$-median inequality is still valid when $\alpha \leq 3.322$. Another key factor in determining the risk of violation is the skewness of the distribution. Previously, it was demonstrated that in a family of distributions differing by a skewness-increasing transformation in van Zwet's sense, the violation of orderliness, if it happens, often only occurs when the distribution is nearly symmetrical (16). The over-corrections in rm and qm are dependent on the SWA_ϵ-median difference, which can be a reasonable measure of skewness (17, 18), implying that the over-correction is often tiny with a moderate d. This qualitative analysis provides another perspective, in addition to the bias bounds (15), that rm and qm based on the mean-SWA_ϵ-median inequality are generally safe.

B. Robust estimations of the central moments. In 1976, Bickel and Lehmann, in their third paper of the landmark series *Descriptive Statistics for Nonparametric Models* (19), generalized a class of estimators called "measures of disperse," which is now often named as Bickel-Lehmann dispersion. As an example, they proposed a first version of the trimmed standard deviation, $\hat{\tau}^2(F;\epsilon) \equiv \tau^2(F;\epsilon)$, for independent and identically distributed random variables X with a distribution F, where $\tau^2(F;\epsilon) = \frac{1}{1-2\epsilon}\int_{Q(\epsilon)}^{Q(1-\epsilon)} y dG(y)$, Q is the quantile function of G, G is the distribution of $Y = X^2$. Obviously, when $\epsilon = 0$, the result is equivalent to the second raw moment. In 1979, in the same series (20), they explored another class of estimators called "measures of spread," which "does not require the assumption of symmetry." From that, a popular efficient scale estimator, the Rousseeuw-Croux scale estimator (21), was derived in 1993, but the importance of tackling the symmetry assumption has been greatly underestimated. In the final section of that paper (20), they considered another two possible versions of the trimmed standard deviations, which were modified here for comparison,

$$\left[n\left(\frac{1}{2}-\epsilon\right)\right]^{-\frac{1}{2}}\left[\sum_{i=\frac{n}{2}}^{n(1-\epsilon)}\left[X_i - X_{n-i+1}\right]^2\right]^{\frac{1}{2}}, \qquad [1]$$

and

$$\left[\binom{n}{2}(1-\epsilon-\gamma\epsilon)\right]^{-\frac{1}{2}}\left[\sum_{i=\binom{n}{2}\epsilon}^{\binom{n}{2}(1-\gamma\epsilon)}\left(X-X'\right)_i^{2}\right]^{\frac{1}{2}}, \qquad [2]$$

where $(X-X')_1 \leq \ldots \leq (X-X')_{\binom{n}{2}}$ are the order statistics of the "pseudo-sample". The paper ended with, "We do not know a fortiori which of the measures [1] or [2] is preferable and leave these interesting questions open."

Observe that the kernel of the unbiased estimation of the second central moment by using U-statistic is $\psi_2(x_1,x_2) = \frac{1}{2}(x_1-x_2)^2$. If adding the $\frac{1}{2}$ term in [2], as $\epsilon \to 0$, the result is equivalent to the standard deviation estimated by using U-statistic (also noted by Janssen, Serfling, and Veraverbeke in 1987) (22). In fact, they also showed that, when ϵ is 0, [2] is $\sqrt{2}$ times the standard deviation.

To address their open question, the nomenclature used in this paper is introduced as follows:

Nomenclature. Given a robust estimator $\hat{\theta}$ with adjustable breakdown point and the breakdown point can be zero. The first part of the name of the robust statistic defined in this paper is a name that indicates the type of estimator, and the second part is the name of the population parameter θ that the estimator is consistent with as $\epsilon \to 0$. The abbreviation of the estimator is formed by combining the initial letter(s) of the first part with the common abbreviation of the consistent estimator that measures the population parameter. If the estimator is symmetric and not a U-statistic, ϵ is indicated in the subscript of the abbreviation of the estimator. For asymmetric estimators, the corresponding γ is also indicated after ϵ. For weighted U-statistics, the breakdown point of the location estimator is indicated, except the median.

In the previous article on semiparametric robust mean estimation, it was shown that the bias of a robust estimator with adjustable breakdown point is often monotonic with respect to the breakdown point in a semiparametric distribution. Naturally, its name should align with the population parameter it is consistent with when $\epsilon \to 0$. The trimmed standard deviation following this nomenclature is $\mathrm{T}sd_{\epsilon,\gamma,n} :=$

$\left[\mathrm{TM}_{\epsilon,\gamma}\left(\left(\psi_2\left(X_{N_1},X_{N_2}\right)\right)_{N=1}^{\binom{n}{2}}\right)\right]^{-\frac{1}{2}}$, where $\mathrm{TM}_{\epsilon,\gamma}(Y)$ denotes the ϵ,γ-trimmed mean with the sequence $\left(\psi_2\left(X_{N_1},X_{N_2}\right)\right)_{N=1}^{\binom{n}{2}}$ as an input. Removing the square root yields the trimmed variance ($\mathrm{T}var_{\epsilon,\gamma,n}$). It is now very clear that this definition, essentially the same as [2], should be preferable. Not only because it is essentially a trimmed U-statistic for the standard deviation but also because the γ-orderliness of the second central moment kernel distribution is ensured by the next exciting theorem.

Theorem B.1. *The second central moment kernel distribution generated from any continuous unimodal distribution is γ-ordered, if $\gamma \leq 1$.*

Proof. Let $Q(p)$, $0 \leq p \leq 1$, denote the quantile of the continuous unimodal distribution $f_X(x)$. The corresponding probability density is $f(Q(p))$. Generating the distribution of the pair $(Q(p_i), Q(p_j))$, $i < j$, $p_i < p_j$, the corresponding probability density is $f_{X,X}(Q(p_i), Q(p_j)) = 2f(Q(p_i))f(Q(p_j))$. Transforming the pair $(Q(p_i), Q(p_j))$, $i < j$, by the function $\Phi(x_1, x_2) = x_1 - x_2$, the pairwise difference distribution has a mode that is arbitrary close to $M - M = 0$. The monotonic increasing of the pairwise difference distribution was first implied in its unimodality proof done by Hodges and Lehmann in 1954 (23). Whereas they used induction to get the result, Dharmadhikari and Jogdeo in 1982 (24) provided a modern proof of the unimodality using Khintchine's representation (25). Assuming absolute continuity, Purkayastha (26) introduced a much simpler proof in 1998. Transforming the pairwise difference distribution by squaring and multiplying by $\frac{1}{2}$ does not change the monotonicity, making the pdf become monotonically decreasing with mode at zero. In the previous article, it was proven that a right skewed distribution with a monotonic decreasing pdf is always γ-ordered, if $\gamma \leq 1$, which gives the desired result. $\square$

Remark. The assumption of continuity of distributions is important for monotonicity because, unlike in the continuous case, it is possible to obtain pairs with the same value for a discrete distribution. For example, let the probabilities of the singletons $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$ and $\{5\}$ of a probability mass function of a discrete probability distribution be $\frac{1}{11}$, $\frac{4}{11}$, $\frac{3}{11}$, $\frac{2}{11}$, and $\frac{1}{11}$, respectively. This is a unimodal distribution, but the corresponding ψ_2 distribution is non-monotonic, whose singletons $\{0\}$, $\{0.5\}$, $\{2\}$, $\{4.5\}$ and $\{8\}$ have probabilities $\frac{21}{66}$, $\frac{24}{66}$, $\frac{2}{14}$, $\frac{6}{66}$, and $\frac{1}{66}$, respectively.

Previously, it was shown that any symmetric distribution with a finite second moment is νth ordered, indicating that orderliness does not require unimodality, e.g., a symmetric bimodal distribution is also ordered. On the other hand, an analysis of the Weibull distribution shows that unimodality does not guarantee orderliness. Theorem B.1 reveals another profound relationship between unimodality and orderliness, which is sufficient for trimming inequality.

In 1928, Fisher constructed k-statistics as unbiased estimators of cumulants (27). Halmos (1946) proved that the functional θ admits an unbiased estimator if and only if it is a regular statistical functional of degree k and showed a relation of symmetry, unbiasness and minimum variance (28). In 1948, Hoeffding generalized U-statistics (29) which enable the derivation of a minimum-variance unbiased estimator from each unbiased estimator of an estimable parameter. Heffernan (1997) (30) obtained an unbiased estimator of the kth central moment by using U-statistics and demonstrated that it is the minimum variance unbiased estimator for distributions with finite moments (31, 32). In 1976, Saleh generalized the Hodges-Lehmenn (H-L) estimator (33) to the trimmed H-L mean (which he named "Wilcoxon one-sample statistic") (34). In 1984, Serfling pointed out the speciality of Hodges-Lehmann estimator, which is neither a simple L-statistic nor a U-statistic, and considered the generalized L-statistics and U-statistic structure (35). Also in 1984, Janssen and Serfling and Veraverbeke (36) showed that the Bickel-Lehmann spread also belongs to the same class. It gradually became clear that the Hodges-Lehmenn estimator, trimmed H-L mean and trimmed standard deviation are all trimmed U-statistics (37–39).

Extending the trimmed U-statistic to weighted U-statistic, i.e., replacing the trimmed mean with weighted average. The weighted kth central moment ($k \leq n$) is defined as,

$$\mathrm{W}km_{\epsilon,\gamma,n} := \mathrm{WA}_{\epsilon,\gamma,n}\left(\left(\psi_k\left(X_{N_1},\cdots,X_{N_k}\right)\right)_{N=1}^{\binom{n}{k}}\right),$$

where $X_{N_1},\cdots,X_{N_k}$ are the n choose k elements from X, $\psi_k\left(x_1,\cdots,x_k\right) = \sum_{j=0}^{k-2}(-1)^j\left(\frac{1}{k-j}\right)\sum\left(x_{i_1}^{k-j}\ldots x_{i_{(j+1)}}\right) + (-1)^{k-1}(k-1)\,x_1\ldots x_k$, the second summation is over $i_1,\cdots,i_{j+1} = 1$ to k with $i_1 < \cdots < i_{j+1}$ (30). Despite the complexity, the structure of the kth central moment kernel distributions can be elucidated by decomposing.

Theorem B.2. *For each pair $\left(Q\left(p_i\right), Q\left(p_j\right)\right)$ of the original distribution, let $x_1 = Q(p_i)$ and $x_k = Q(p_j)$, $\Delta = Q(p_i) - Q(p_j)$. The kth central moment kernel distribution, $k > 2$, can be seen as a mixture distribution and each of the components has the support $\left(-\left(\frac{k}{\frac{3+(-1)^k}{2}}\right)^{-1}(-\Delta)^k, \frac{1}{k}(-\Delta)^k\right)$.*

Proof. Generating the distribution of the k-tuple $(Q(p_{i_1}),\ldots,Q(p_{i_k}))$, $k > 2$, $i_1 < \ldots < i_k$, $p_{i_1} < \ldots < p_{i_k}$, the corresponding probability density is $f_{X,\ldots,X}(Q(p_{i_1}),\ldots,Q(p_{i_k})) = k!f(Q(p_{i_1}))\ldots f(Q(p_{i_k}))$. Transforming the distribution of the k-tuple by the function $\psi_k\left(x_1,\ldots,x_k\right)$, denoting $\bar{\Delta} = \psi_k\left(Q(p_{i_1}),\ldots,Q(p_{i_k})\right)$. The probability $f_{\Xi_k}(\bar{\Delta}) = \sum_{\bar{\Delta}=\psi_k\left(Q(p_{i_1}),\ldots,Q(p_{i_k})\right)} f_{X,\ldots,X}(Q(p_{i_1}),\ldots,Q(p_{i_k}))$ is the summation of the probabilities of all k-tuples such that $\bar{\Delta}$ is equal to $\psi_k\left(Q(p_{i_1}),\ldots,Q(p_{i_k})\right)$. The following Ξ_k is equivalent.

Ξ_k: Every pair with a difference equal to $\Delta = Q(p_{i_1}) - Q(p_{i_k})$ can generate a pseudodistribution (but the integral is not equal to 1, so "pseudo") such that x_2, …, x_{k-1} exhaust all combinations under the inequality constraints, i.e., $Q(p_{i_1}) = x_1 < x_2 < \ldots < x_{k-1} < x_k = Q(p_{i_k})$. The combination of all the pseudodistributions with the same Δ is ξ_Δ. The combination of ξ_Δ, i.e., from $\Delta = 0$ to $Q(0) - Q(1)$, is Ξ_k.

The support of ξ_Δ is the extrema of ψ_k subject to the inequality constraints. Using the Lagrange multiplier, one can easily determine the only critical point at $x_1 = \ldots = x_k = 0$, where $\psi_k = 0$. Other candidates are within the boundaries, i.e., $\psi_k\left(x_1 = x_1, x_2 = x_k, \cdots, x_k = x_k\right)$, $\cdots$, $\psi_k\left(x_1 = x_1, \cdots, x_i = x_1, x_{i+1} = x_k, \cdots, x_k = x_k\right)$, $\cdots$, $\psi_k\left(x_1 = x_1, \cdots, x_{k-1} = x_1, x_k = x_k\right)$. $\psi_k\left(x_1 = x_1, \cdots, x_i = x_1, x_{i+1} = x_k, \cdots, x_k = x_k\right)$ can be divided into k groups. If $\frac{k+1-i}{2} \leq j \leq \frac{k-1}{2}$, from $j+1$st to $k-j$th group, the gth group has $i\binom{i-1}{g-j-1}\binom{k-i}{j}$ terms having the form $(-1)^{g+1}\frac{1}{k-g+1}x_1^{k-j}x_k^{j}$, from $k-j+1$th to $i+j$th group, the gth group has $i\binom{i-1}{g-j-1}\binom{k-i}{j}+(k-i)\binom{k-i-1}{j-k+g-1}\binom{i}{k-j}$ terms having the form $(-1)^{g+1}\frac{1}{k-g+1}x_1^{k-j}x_k^{j}$. If $j<\frac{k+1-i}{2}$, from $j+1$st to $i+j$th group, the gth group has $i\binom{i-1}{g-j-1}\binom{k-i}{j}$ terms having the form $(-1)^{g+1}\frac{1}{k-g+1}x_1^{k-j}x_k^{j}$. If $j\geq\frac{k}{2}$, from $k-j+1$st to jth group, the gth group has $(k-i)\binom{k-i-1}{j-k+g-1}\binom{i}{k-j}$ terms having the form $(-1)^{g+1}\frac{1}{k-g+1}x_1^{k-j}x_k^{j}$, from $j+1$th to $j+i$th group, $i+j<k$, the gth group has $i\binom{i-1}{g-j-1}\binom{k-i}{j}+(k-i)\binom{k-i-1}{j-k+g-1}\binom{i}{k-j}$ terms having the form $(-1)^{g+1}\frac{1}{k-g+1}x_1^{k-j}x_k^{j}$. The final kth group is the term $(-1)^{k-1}(k-1)x_1^{i}x_k^{k-i}$. So, if $i+j=k$, $j\geq\frac{k}{2}$, $i\leq\frac{k}{2}$, the summed coefficient of $x_1^{i}x_k^{k-i}$ is $(-1)^{k-1}(k-1)+\sum_{g=i+1}^{k-1}(-1)^{g+1}\frac{1}{k-g+1}(k-i)\binom{k-i-1}{g-i-1}+\sum_{g=k-i+1}^{k-1}(-1)^{g+1}\frac{1}{k-g+1}i\binom{i-1}{g-k+i-1}=(-1)^{k-1}(k-1)+(-1)^{k+1}+(k-i)(-1)^{k}+(-1)^{k}(i-1)=(-1)^{k+1}$. The summation identities are $\sum_{g=i+1}^{k-1}(-1)^{g+1}\frac{1}{k-g+1}(k-i)\binom{k-i-1}{g-i-1}=(k-i)\int_0^1\sum_{g=i+1}^{k-1}(-1)^{g+1}\binom{k-i-1}{g-i-1}t^{k-g}dt=(k-i)\int_0^1\left((-1)^{i}(t-1)^{k-i-1}-(-1)^{k+1}\right)dt=(k-i)\left(\frac{(-1)^{k}}{i-k}+(-1)^{k}\right)=(-1)^{k+1}+(k-i)(-1)^{k}$ and $\sum_{g=k-i+1}^{k-1}(-1)^{g+1}\frac{1}{k-g+1}i\binom{i-1}{g-k+i-1}=\int_0^1\sum_{g=k-i+1}^{k-1}(-1)^{g+1}i\binom{i-1}{g-k+i-1}t^{k-g}dt=\int_0^1\left(i(-1)^{k-i}(t-1)^{i-1}-i(-1)^{k+1}\right)dt=(-1)^{k}(i-1)$. If $j<\frac{k+1-i}{2}$, $i>k-1$, if $i=k$, $\psi_k=0$, if $\frac{k+1-i}{2}\leq j\leq\frac{k-1}{2}$, $\frac{k+1}{2}\leq i\leq k-1$, the summed coefficient of $x_1^{i}x_k^{k-i}$ is $(-1)^{k-1}(k-1)+\sum_{g=k-i+1}^{k-1}(-1)^{g+1}\frac{1}{k-g+1}i\binom{i-1}{g-k+i-1}+\sum_{g=i+1}^{k-1}(-1)^{g+1}\frac{1}{k-g+1}(k-i)\binom{k-i-1}{g-i-1}$, the same as above. If $i+j<k$, since $\binom{i}{k-j}=0$, the related terms can be ignored, so, using the binomial theorem and beta function, the summed coefficient of $x_1^{k-j}x_k^{j}$ is $\sum_{g=j+1}^{i+j}(-1)^{g+1}\frac{1}{k-g+1}i\binom{i-1}{g-j-1}\binom{k-i}{j}=i\binom{k-i}{j}\int_0^1\sum_{g=j+1}^{i+j}(-1)^{g+1}\binom{i-1}{g-j-1}t^{k-g}dt=\binom{k-i}{j}i\int_0^1\left((-1)^{j}t^{k-j-1}\left(\frac{t}{t-1}\right)^{1-i}\right)dt=\binom{k-i}{j}i\frac{(-1)^{j+i+1}\Gamma(i)\Gamma(k-j-i+1)}{\Gamma(k-j+1)}=\frac{(-1)^{j+i+1}i!(k-j-i)!(k-i)!}{(k-j)!j!(k-j-i)!}=(-1)^{j+i+1}\frac{i!(k-i)!}{k!}\frac{k!}{(k-j)!j!}=\binom{k}{i}^{-1}(-1)^{1+i}\binom{k}{j}(-1)^{j}$.

The coefficient of $x_1^{i}x_k^{k-i}$ in $\binom{k}{i}^{-1}(-1)^{1+i}(x_1-x_k)^{k}$ is $\binom{k}{i}^{-1}(-1)^{1+i}\binom{k}{i}(-1)^{k-i}=(-1)^{k+1}$, same as the summed coefficient if $i+j=k$. If $i+j<k$, the coefficient of $x_1^{k-j}x_k^{j}$ is $\binom{k}{i}^{-1}(-1)^{1+i}\binom{k}{j}(-1)^{j}$, same as the corresponding summed coefficient. Therefore, $\psi_k(x_1=x_1,\ldots,x_i=x_1,x_{i+1}=x_k,\ldots,x_k=x_k)=\binom{k}{i}^{-1}(-1)^{1+i}(x_1-x_k)^{k}$, the maximum and minimum of ψ_k follow directly from the properties of the binomial coefficient. □

ξ_Δ is closely related to $f_\Xi(\Delta)$, which is the pairwise difference distribution, since the probability density of ξ_Δ can be expressed as $f_{\Xi_k}(\bar{\Delta}|\Delta)$ and $\sum_{\bar{\Delta}=-\left(\binom{k}{\frac{3+(-1)^k}{2}}\right)^{-1}(-\Delta)^k}^{\frac{1}{k}(-\Delta)^k} f_{\Xi_k}(\bar{\Delta}|\Delta)=f_\Xi(\Delta)$. Recall that $f_\Xi(\Delta)$ is monotonic increasing with a mode at the origin if the original distribution is unimodal. Thus, in general, ignoring the shape of ξ_Δ, Ξ_k is monotonic left and right around zero. In fact, the median of Ξ_k also exhibits a strong tendency to be close to zero, as it can be cast as a weighted mean of the medians of ξ_Δ. When Δ is small, all values of ξ_Δ are close to zero, resulting in the median of ξ_Δ being close to zero as well. When Δ is large, the median of ξ_Δ depends on its skewness, but the corresponding weight is much smaller, so even if ξ_Δ is highly skewed, the median of Ξ_k will only be slightly shifted from zero. Denote the median of Ξ_k as m_{Ξ_k}, for the five parametric distributions here, $|m_{\Xi_k}|$s are all $\leq 0.1\sigma$ for Ξ_3 and Ξ_4 (SI Dataset S1). Assuming $m_{\Xi_k}=0$, for the even ordinal central moment kernel distribution, the average probability density on the left side of zero is greater than that on the right side, since $\frac{\frac{1}{2}}{\binom{k}{2}^{-1}(Q(0)-Q(1))^k}>\frac{\frac{1}{2}}{\frac{1}{k}(Q(0)-Q(1))^k}$. This means that, on average, the inequality $f(Q(\epsilon))\geq f(Q(1-\epsilon))$ holds. For the odd ordinal distribution, the discussion is more challenging since it is generally symmetric. Just consider Ξ_3, let $x_1=Q(p_i)$ and $x_3=Q(p_j)$, changing the value of x_2 from $Q(p_i)$ to $Q(p_j)$ will monotonically change the value of $\psi_3(x_1,x_2,x_3)$, since $\frac{\partial\psi_3(x_1,x_2,x_3)}{\partial x_2}=-\frac{x_1^2}{2}-x_1x_2+2x_1x_3+x_2^2-x_2x_3-\frac{x_3^2}{2}$, $-\frac{3}{4}(x_1-x_3)^2\leq\frac{\partial\psi_3(x_1,x_2,x_3)}{\partial x_2}\leq-\frac{1}{2}(x_1-x_3)^2\leq 0$. If the original distribution is right-skewed, ξ_Δ will be left-skewed, so, for Ξ_3, the average probability density of the right side of zero will be greater than that of the left side, which means, on average, the inequality $f(Q(\epsilon))\leq f(Q(1-\epsilon))$ holds (the same result can be inferred from the definition of central moments, where the positivity of the odd order central moment is directly related to the left-skewness of the corresponding kernel distribution). In all, the monotonicity of the pairwise difference distribution guides the general shape of the kth central moment kernel distribution, $k>2$, forcing it to be unimodal-like with the mode and median close to zero, then, the inequality $f(Q(\epsilon))\leq f(Q(1-\epsilon))$ or $f(Q(\epsilon))\geq f(Q(1-\epsilon))$ holds in general. If a distribution is ordered and all of its central moment kernel distributions are also ordered, it is called completely ordered. Although strict complete orderliness is difficult to prove, even if the inequality may be violated in a small range, as discussed in Subsection A, the mean-SWA$_\epsilon$-median inequality remains valid, in most cases, for the central moment kernel distribution.

Another crucial property of the central moment kernel distribution, location invariant, is introduced in the next theorem. The proof is provided in the SI Text.

Theorem B.3. $\psi_k(x_1=\lambda x_1+\mu,\cdots,x_k=\lambda x_k+\mu)=\lambda^k\psi_k(x_1,\cdots,x_k)$.

Consider two continuous distributions belonging to the same location–scale family, their corresponding kth central moment kernel distributions only differ in scaling. So d is invariant, as shown in Subsection A. The recombined kth central moment, based on rm, is defined by,

$$rkm_{d,\epsilon,n}\coloneqq(d+1)\,\mathrm{SW}km_{\epsilon,n}-dmkm_n,$$

where $\mathrm{SW}km_{\epsilon,n}$ is using the binomial kth central moment ($\mathrm{B}km_{\epsilon,n}$) here, mkm_n is the median kth central moment. Since $\mathrm{SW}km_{\epsilon,n}$ is an L-statistic, the resulting $rkm_{d,\epsilon,n}$ is an arithmetic I-statistic. Similarly, the quantile will not change after scaling. The quantile kth central moment is thus defined as

$$qkm_{d,\epsilon,n}\coloneqq\hat{Q}_n\left(\left(p\mathrm{SW}km_{\epsilon,n}-\frac{1}{2}\right)d+p\mathrm{SW}km_{\epsilon,n}\right),$$

where $pSWkm_{\epsilon,n} = \hat{F}_{\psi,n}(SWkm_{\epsilon,n})$, $\hat{F}_{\psi,n}$ is the empirical cumulative distribution function of the corresponding central moment kernel distribution. $qkm_{d,\epsilon,n}$ is a quantile I-statistic.

Finally, for standardized moments, quantile skewness and quantile kurtosis are defined to be $qskew_{d,\epsilon,n} := \frac{qtm_{d,\epsilon,n}}{qsd^3_{d,\epsilon,n}}$ and $qkurt_{d,\epsilon,n} := \frac{qfm_{d,\epsilon,n}}{qsd^4_{d,\epsilon,n}}$. Quantile standard deviation ($qsd_{d,\epsilon,n}$), recombined standard deviation ($rsd_{d,\epsilon,n}$), quantile third central moment ($qtm_{d,\epsilon,n}$), quantile fourth central moment ($qfm_{d,\epsilon,n}$), recombined third central moment ($rtm_{d,\epsilon,n}$), recombined fourth central moment ($rfm_{d,\epsilon,n}$), recombined skewness ($rskew_{d,\epsilon,n}$), and recombined kurtosis ($rkurt_{d,\epsilon,n}$) are all defined similarly as above and not repeated here. The transformation to a location problem can also empower related statistical tests. From the better performance of the quantile mean in heavy-tailed distributions, quantile central moments are generally better than recombined central moments regarding asymptotic bias.

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

ACKNOWLEDGMENTS. I gratefully acknowledge the constructive comments made by the editor which substantially improved the clarity and quality of this paper.

1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae.* (Henricus Dieterich), (1823).
2. S Newcomb, A generalized theory of the combination of observations so as to obtain the best result. *Am. journal Math.* **8**, 343–366 (1886).
3. S Newcomb, Researches on the motion of the moon. part ii, the mean motion of the moon and other astronomical elements derived from observations of eclipses and occultations extending from the period of the babylonians until ad 1908. *United States. Naut. Alm. Off. Astron. paper; v. 9* **9**, 1 (1912).
4. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
5. X He, WK Fung, Method of medians for lifetime data with weibull models. *Stat. medicine* **18**, 1993–2009 (1999).
6. M Menon, Estimation of the shape and scale parameters of the weibull distribution. *Technometrics* **5**, 175–182 (1963).
7. SD Dubey, Some percentile estimators for weibull parameters. *Technometrics* **9**, 119–129 (1967).
8. KM Hassanein, Percentile estimators for the parameters of the weibull distribution. *Biometrika* **58**, 673–676 (1971).
9. NB Marks, Estimation of weibull parameters from common percentiles. *J. applied Stat.* **32**, 17–24 (2005).
10. K Boudt, D Caliskan, C Croux, Robust explicit estimators of weibull parameters. *Metrika* **73**, 187–209 (2011).
11. SD Dubey, *Contributions to statistical theory of life testing and reliability.* (Michigan State University of Agriculture and Applied Science. Department of statistics), (1960).
12. LJ Bain, CE Antle, Estimation of parameters in the weibdl distribution. *Technometrics* **9**, 621–627 (1967).
13. RV Hogg, Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *J. Am. Stat. Assoc.* **69**, 909–923 (1974).
14. RJ Hyndman, Y Fan, Sample quantiles in statistical packages. *The Am. Stat.* **50**, 361–365 (1996).
15. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
16. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
17. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
18. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).
19. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models. iii. dispersion in *Selected works of EL Lehmann.* (Springer), pp. 499–518 (2012).
20. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models iv. spread in *Selected Works of EL Lehmann.* (Springer), pp. 519–526 (2012).
21. PJ Rousseeuw, C Croux, Alternatives to the median absolute deviation. *J. Am. Stat. association* **88**, 1273–1283 (1993).
22. P Janssen, R Serfling, N Veraverbeke, Asymptotic normality of u-statistics based on trimmed samples. *J. statistical planning inference* **16**, 63–74 (1987).
23. J Hodges, E Lehmann, Matching in paired comparisons. *The Annals Math. Stat.* **25**, 787–791 (1954).
24. S Dharmadhikari, K Jogdeo, Unimodal laws and related in *A Festschrift For Erich L. Lehmann.* (CRC Press), p. 131 (1982).
25. AY Khintchine, On unimodal distributions. *Izv. Nauchno-Isled. Inst. Mat. Mech.* **2**, 1–7 (1938).
26. S Purkayastha, Simple proofs of two results on convolutions of unimodal distributions. *Stat. & probability letters* **39**, 97–100 (1998).
27. RA Fisher, Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.* **2**, 199–238 (1930).
28. PR Halmos, The theory of unbiased estimation. *The Annals Math. Stat.* **17**, 34–43 (1946).
29. W Hoeffding, A class of statistics with asymptotically normal distribution. *The Annals Math. Stat.* **19**, 293–325 (1948).
30. PM Heffernan, Unbiased estimation of central moments by using u-statistics. *J. Royal Stat. Soc. Ser. B (Statistical Methodol.* **59**, 861–863 (1997).
31. D Fraser, Completeness of order statistics. *Can. J. Math.* **6**, 42–45 (1954).
32. AJ Lee, *U-statistics: Theory and Practice.* (Routledge), (2019).
33. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
34. A Ehsanes Saleh, Hodges-lehmann estimate of the location parameter in censored samples. *Annals Inst. Stat. Math.* **28**, 235–247 (1976).
35. RJ Serfling, Generalized l-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).
36. P Janssen, R Serfling, N Veraverbeke, Asymptotic normality for a general class of statistical functions and applications to measures of spread. *The Annals Stat.* **12**, 1369–1379 (1984).
37. MG Akritas, Empirical processes associated with v-statistics and a class of estimators under random censoring. *The Annals Stat.* **14**, 619–637 (1986).
38. I Gijbels, P Janssen, N Veraverbeke, Weak and strong representations for trimmed u-statistics. *Probab. theory related fields* **77**, 179–194 (1988).
39. J Choudhury, R Serfling, Generalized order statistics, bahadur representations, and sequential nonparametric fixed-width confidence intervals. *J. Stat. Plan. Inference* **19**, 269–282 (1988).