

Near-consistent robust estimations of moments for unimodal distributions

Tuban Lee

This manuscript was compiled on May 22, 2023

Descriptive statistics for parametric models currently heavily rely on the accuracy of distributional assumptions. Here, leveraging 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 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 potential inconsistencies between the sample mean (\bar{x}) and robust location estimators with non-zero asymptotic breakdown points in distributions with finite moments 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) provided the first modern approach to robust parametric estimation by developing a class of estimators that gives "less weight to the more discordant observations" (2). In 1964, Huber (3) 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 parameter estimations. For example, He and Fung (1999) constructed (4) 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 interesting approach is based on L -estimators, such as percentile estimators. Examples of percentile estimators for the Weibull distribution, the reader is referred to Menon (1963) (5), Dubey (1967) (6), Marks (2005) (7), and Boudt, Caliskan, and Croux (2011) (8)'s works. At the outset of the study of percentile estimators, it was known that they arithmetically utilize the invariant structures of probability distributions (5, 6). Maybe such estimators can be named as I -statistics. Formally, an estimator is classified as an I -statistic if it asymptotically satisfies $I(LE_1, \dots, LE_l) = (\theta_1, \dots, \theta_q)$ for the distribution it is consistent, where LE s are calculated with the use of LU -statistics (defined in Subsection B), I is defined using arithmetic operations and constants but may also incorporate transcendental functions and quantile functions, and θ s are the population parameters it estimates. A subclass of I -statistics, arithmetic I -statistics, is defined as LE s are LU -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 -estimators, a percentile estimator might not always be an arithmetic I -statistic (6). 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 LU -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 -estimator can be expressed as an integral of the quantile function, I -statistics are often analytically obtainable. However, it is observed that Marks percentile estimator for the Weibull distribution (7) tends to be inferior to the robust M -estimators (3, 4), especially upon violation of the distributional assumption (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 for central moments, rendering the approach ill-suited (SI Dataset S1).

The majority of robust location estimators commonly used are symmetric owing to the prevalence of symmetric distributions. An asymmetric weighted L -statistic can achieve consistency for a semiparametric class of skewed distributions; but the lack of symmetry makes it suitable only for certain applications. Shifting from semiparametrics to parametrics, consider an estimator with a non-zero asymptotic breakdown point that is consistent simultaneously for both a semiparametric class of distributions and a distinct parametric distribution with finite moments, such a robust location estimator is called an invariant mean. Based on the mean-weighted L -statistic- γ -median inequality, the recombined mean is defined as

$$rm_{d,\epsilon,\gamma,n} := \lim_{c \rightarrow \infty} \left(\frac{(WL_{\epsilon,\gamma,n} + c)^{d+1}}{(\gamma m_n + c)^d} - c \right),$$

where d is the key factor for bias correction, γm_n is the sample γ -median, $WL_{\epsilon,\gamma,n}$ is the weighted L -statistic. If γ is omitted, $\gamma = 1$ is assumed. The subsequent theorem shows the significance of this arithmetic I -statistic.

Significance Statement

Bias, variance, and contamination are the three main errors in statistics. Consistent robust estimation is unattainable without parametric assumptions. In this article, 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.

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

Theorem .1. Let $BM_{\epsilon,n}$ be the WL, $rm_{d \approx 0.103, \epsilon = \frac{1}{24}}$ is a consistent mean estimator for the exponential distribution, any symmetric distributions and the Pareto distribution with quantile function $Q(p) = x_m(1-p)^{-\frac{1}{\alpha}}$, $x_m > 0$, when $\alpha \rightarrow \infty$, provided that the second moments are finite.

Proof. Finding d and ϵ that make $rm_{d,\epsilon}$ a consistent mean estimator is equivalent to finding the solution of $E[rm_{d,\epsilon,n}] = E[X]$. The quantile function of the exponential distribution is $Q(p) = \ln\left(\frac{1}{1-p}\right)\lambda$. $E[X] = \lambda$. $E[m_n] = Q\left(\frac{1}{2}\right) = \ln 2\lambda$. For the exponential distribution, $E\left[BM_{\epsilon=\frac{1}{24},n}\right] = \lambda\left(1 + \ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)$, the detailed formula is given in the SI Text. Since $rm_{d,\epsilon} = \lim_{c \rightarrow \infty} \left(\frac{(BM_{\epsilon}+c)^{d+1}}{(m+c)^d} - c\right) = (d+1)BM_{\epsilon} - dm = \mu$. So,

$$d = \frac{\mu - BM_{\epsilon}}{BM_{\epsilon} - m} = \frac{\lambda - \lambda\left(1 + \ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)}{\lambda\left(1 + \ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right) - \ln 2\lambda} =$$

$$-\frac{\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)}{1 - \ln(2) + \ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)} \approx 0.103.$$

The proof of the second assertion follows directly from the coincidence property. For any symmetric distribution with a finite second moment, $E[BM_{\epsilon,n}] = E[m_n] = E[X]$. Then $E[rm_{d,\epsilon,n}] = \lim_{c \rightarrow \infty} \left(\frac{(E[X]+c)^{d+1}}{(E[X]+c)^d} - c\right) = E[X]$. The proof for the Pareto distribution is more general. The mean of the Pareto distribution is given by $\frac{\alpha x_m}{\alpha-1}$. Since any weighted L -statistic can be expressed as an integral of the quantile function as shown in Theorem A.1, the γ -median is also a percentile, replacing the WL and γm in the d value with two arbitrary percentiles p_1 and p_2 , for the Pareto distribution,

$$d_{Pareto} = \frac{\mu - Q(p_1)}{Q(p_1) - Q(p_2)} = \frac{\frac{\alpha x_m}{\alpha-1} - x_m(1-p_1)^{-\frac{1}{\alpha}}}{x_m(1-p_1)^{-\frac{1}{\alpha}} - x_m(1-p_2)^{-\frac{1}{\alpha}}}.$$

x_m can be canceled out. For the exponential distribution, $d_{exp} = \frac{\mu - Q(p_1)}{Q(p_1) - Q(p_2)} = \frac{\lambda - \ln\left(\frac{1}{1-p_1}\right)\lambda}{\ln\left(\frac{1}{1-p_1}\right)\lambda - \ln\left(\frac{1}{1-p_2}\right)\lambda} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}.$

Since $\lim_{\alpha \rightarrow \infty} \frac{\frac{\alpha}{\alpha-1} - (1-p_1)^{-1/\alpha}}{(1-p_1)^{-1/\alpha} - (1-p_2)^{-1/\alpha}} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$, the d value for the Pareto distribution approaches that of the exponential distribution, as $\alpha \rightarrow \infty$, regardless of the type of weighted L -statistic used. This completes the demonstration. \square

Theorem .1 implies that for the Weibull, gamma, Pareto, lognormal and generalized Gaussian distribution, $rm_{d \approx 0.103, \epsilon = \frac{1}{24}}$ is consistent for at least one particular case. The biases of $rm_{d \approx 0.103, \epsilon = \frac{1}{24}}$ for distributions with skewness between those of the exponential and symmetric distributions are tiny (SI Dataset S1). $rm_{d \approx 0.103, \epsilon = \frac{1}{24}}$ 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 $n = 4096$) for unimodal distributions.

Background and Main Results

A. Invariant mean. It is well established 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, parametrized by a location parameter μ and a scale parameter λ , takes the form $F(x) = F_0\left(\frac{x-\mu}{\lambda}\right)$, where F_0 is a standard distribution without any shifts or scaling. Consequently, $F(x) = Q^{-1}(x) \rightarrow x = Q(p) = \lambda Q_0(p) + \mu$. Thus, for a location-scale distribution, any $WA(\epsilon, \gamma)$ can be expressed as $\lambda WA_0(\epsilon, \gamma) + \mu$, where $WA_0(\epsilon, \gamma)$ is an integral of $Q_0(p)$ according to the definition of the weighted average. 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. The proof is given in the SI Text.

Theorem A.1. $whl_k(x_1 = \lambda x_1 + \mu, \dots, x_k = \lambda x_k + \mu) = \lambda whl_k(x_1, \dots, x_k) + \mu$.

Let $WeHLM_0(\epsilon, \gamma)$ denote the expected value of a weighted Hodges-Lehmann mean for the standard distribution, then for the same location-scale distribution above, the $WeHLM$ can also be expressed as $\lambda WeHLM_0(\epsilon, \gamma) + \mu$. As Theorem A.1 proved the $w_i \neq 1$ case, this form is valid for all weighted L -statistics. The simultaneous cancellation of μ and λ in $\frac{(\lambda\mu_0 + \mu) - (\lambda WL_0(\epsilon, \gamma) + \mu)}{(\lambda WL_0(\epsilon, \gamma) + \mu) - (\lambda \gamma m_0 + \mu)}$ assures that d is always a constant for a location-scale distribution.

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

$$qm_{d,\epsilon,\gamma,n} := \hat{Q}_n \left(\left(\hat{F}_n(WL_{\epsilon,\gamma,n}) - \frac{1}{1+\gamma} \right) d + \hat{F}_n(WL_{\epsilon,\gamma,n}) \right),$$

provided that $\hat{F}_n(WL_{\epsilon,\gamma,n}) \geq \frac{1}{1+\gamma}$, where $\hat{F}_n(x)$ is the empirical cumulative distribution function of the sample, \hat{Q}_n is the sample quantile function. When $\hat{F}_n(WL_{\epsilon,\gamma,n}) < \frac{1}{1+\gamma}$, $qm_{d,\epsilon,\gamma,n}$ is defined as $\hat{Q}_n \left(\hat{F}_n(WL_{\epsilon,\gamma,n}) - \left(\frac{1}{1+\gamma} - \hat{F}_n(WL_{\epsilon,\gamma,n}) \right) d \right)$. Without loss of generality, in the following discussion, only the case where $\hat{F}_n(WL_{\epsilon,\gamma,n}) \geq \frac{1}{1+\gamma}$ is considered. Moreover, in extreme right-skewed heavy-tailed distributions, if the calculated percentile surpasses $1 - \epsilon$, it will be modified to $1 - \epsilon$. A widely used method for calculating the sample quantile function involves employing linear interpolation of modes corresponding to the order statistics of the uniform distribution on the interval $[0, 1]$, i.e., $\hat{Q}_n(p) = X_{[h]} + (h - [h])(X_{[h]} - X_{[h]})$, $h = (n-1)p + 1$. To minimize the finite sample bias, here, the inverse function of \hat{Q}_n is deduced as $\hat{F}_n(x) := \frac{1}{n-1} \left(cf - 1 + \frac{x - X_{cf}}{X_{cf+1} - X_{cf}} \right)$, where $cf = \sum_{i=1}^n \mathbf{1}_{X_i \leq x}$, $\mathbf{1}_A$ is the indicator of event A . The quantile mean uses the location-scale invariant in a different way as shown in the subsequent proof.

Theorem A.2. Let $BM_{\epsilon,n}$ be the WL, $qm_{d \approx 0.088, \epsilon = \frac{1}{24}}$ is a consistent mean estimator for the exponential, Pareto ($\alpha \rightarrow \infty$) and any symmetric distributions provided that the second moments are finite.

164 *Proof.* The cdf of the exponential distribution is $F(x) =$
165 $1 - e^{-\lambda^{-1}x}$, $\lambda \geq 0$, $x \geq 0$. Recall that the expecta-
166 tion of $\text{BM}_{\epsilon,n}$ can be expressed as $\lambda \text{BM}_0(\epsilon)$, so $F(\text{BM}_\epsilon)$ is
167 free of λ , as are $F(\mu)$ and $F(m)$. When $\epsilon = \frac{1}{24}$, $d =$
168
$$\frac{F(\mu) - F(\text{BM}_\epsilon)}{F(\text{BM}_\epsilon) - \frac{1}{2}} = \frac{-e^{-1} + e^{-\left(1 + \ln\left(\frac{26068394603446272 \sqrt[6]{\frac{7}{247}} \sqrt[3]{\frac{11}{11}}\right)}\right)}}{-\left(1 + \ln\left(\frac{26068394603446272 \sqrt[6]{\frac{7}{247}} \sqrt[3]{\frac{11}{11}}\right)}\right)}} =$$

169
$$\frac{\frac{1}{2} - e^{-\left(1 + \ln\left(\frac{26068394603446272 \sqrt[6]{\frac{7}{247}} \sqrt[3]{\frac{11}{11}}\right)}\right)}}{\frac{1}{2} - e^{-\left(1 + \ln\left(\frac{26068394603446272 \sqrt[6]{\frac{7}{247}} \sqrt[3]{\frac{11}{11}}\right)}\right)}} \approx 0.088.$$
 The proof of the
170 symmetric case: since for any symmetric distribution with
171 a finite second moment, $F(E[\text{BM}_{\epsilon,n}]) = F(\mu) = \frac{1}{2}$.
172 Then, the expectation of the quantile mean is $qm_{d,\epsilon} =$
173 $F^{-1}\left(\left(F(\mu) - \frac{1}{2}\right)d + F(\mu)\right) = F^{-1}\left(0 + F(\mu)\right) = \mu$.
174 For the assertion related to the Pareto distribu-
175 tion, the cdf of it is $1 - \left(\frac{x_m}{x}\right)^\alpha$. Similar to The-
176 orem .1, replacing the $F(\text{WL}_{\epsilon,\gamma})$ and $\frac{1}{1+\gamma}$ in the
177 d value with two arbitrary percentiles p_1 and p_2 ,
178
$$d_{\text{Pareto}} = \frac{1 - \left(\frac{x_m}{x_m(1-p_1)}\right)^\alpha - \left(1 - \left(\frac{x_m}{x_m(1-p_2)}\right)^\alpha\right)}{\left(1 - \left(\frac{x_m}{x_m(1-p_1)}\right)^\alpha\right) - \left(1 - \left(\frac{x_m}{x_m(1-p_2)}\right)^\alpha\right)} =$$

179
$$\frac{1 - \left(\frac{\alpha-1}{\alpha}\right)^\alpha - p_1}{p_1 - p_2}.$$
 When $\alpha \rightarrow \infty$, $\left(\frac{\alpha-1}{\alpha}\right)^\alpha = \frac{1}{e}$, so in this
180 case, d_{Pareto} is identical to that of the exponential distri-
181 bution, since $d_{\text{exp}} = \frac{(1-e^{-1}) - \left(1 - e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)}{\left(1 - e^{-\ln\left(\frac{1}{1-p_1}\right)}\right) - \left(1 - e^{-\ln\left(\frac{1}{1-p_2}\right)}\right)} =$
182
$$\frac{1 - \frac{1}{e} - p_1}{p_1 - p_2}.$$
 Therefore, same logic as in Theorem .1, their d val-
183 ues are always identical, regardless of the type of weighted
184 L -statistic used. All results are now proven. \square

185 The definitions of location and scale parameters are such
186 that they must satisfy $F(x; \lambda, \mu) = F\left(\frac{x-\mu}{\lambda}; 1, 0\right)$. By recalling
187 $x = \lambda Q_0(p) + \mu$, it follows that the percentile of any weighted
188 L -statistic is free of λ and μ , which guarantees the validity
189 of the quantile mean. The quantile mean is a quantile I -
190 statistic. Specifically, an estimator is classified as a quantile
191 I -statistic if LEs are percentiles of a distribution obtained by
192 plugging LU -statistics into a cumulative distribution function
193 and I is defined with arithmetic operations, constants and
194 quantile functions. $qm_{d \approx 0.088, \epsilon = \frac{1}{24}}$ works better in the fat-tail
195 scenarios (SI Dataset S1). Theorem .1 and A.2 show that
196 $rm_{d \approx 0.103, \epsilon = \frac{1}{24}}$ and $qm_{d \approx 0.088, \epsilon = \frac{1}{24}}$ are both consistent mean
197 estimators for any symmetric distribution and a skewed dis-
198 tribution with finite second moments. The breakdown points
199 of $rm_{d \approx 0.103, \epsilon = \frac{1}{24}}$ and $qm_{d \approx 0.088, \epsilon = \frac{1}{24}}$ are both $\frac{1}{24}$. Therefore
200 they are all invariant means.

201 To study the impact of the choice of WLs in rm and qm ,
202 it is constructive to recall that a weighted L -statistic is a
203 combination of order statistics. While using a less-biased
204 weighted L -statistic can generally enhance performance (SI
205 Dataset S1), there is a greater risk of violation in the semi-
206 parametric framework. However, the mean- $\text{WA}_{\epsilon,\gamma}$ -median
207 inequality is robust to slight fluctuations of the QA function
208 of the underlying distribution when $0 \leq \gamma \leq 1$. Suppose the
209 QA function is generally decreasing in $[0, u]$, but increasing
210 in $[u, \frac{1}{1+\gamma}]$, since all quantile averages with breakdown points

from ϵ to $\frac{1}{1+\gamma}$ will be included in the computation of $\text{WA}_{\epsilon,\gamma}$,
as long as $\frac{1}{1+\gamma} - u \ll \frac{1}{1+\gamma} - \gamma\epsilon$, and other portions of the QA
function satisfy the inequality constraints that define the ν th
 γ -orderliness on which the $\text{WA}_{\epsilon,\gamma}$ is based, the mean- $\text{WA}_{\epsilon,\gamma}$ -
 γ -median inequality still holds. This is due to the violation
being bounded (9) when $0 \leq \gamma \leq 1$ 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 pre-
vious article, the violation of the third orderliness starts near
this parameter as well, yet the mean-BM $\frac{1}{24}$ -median inequality
retains valid when $\alpha \leq 3.387$. Another key factor in determin-
ing the risk of violation is the skewness of the distribution.
Previously, it was demonstrated that in a family of distribu-
tions differing by a skewness-increasing transformation in van
Zwet's sense, the violation of orderliness, if it happens, often
only occurs as the distribution nears symmetry (10). When
 $\gamma = 1$, the over-corrections in rm and qm are dependent on the
 SWA_ϵ -median difference, which can be a reasonable measure
of skewness after standardization (11, 12), implying that the
over-correction is often tiny with moderate d . This qualitative
analysis suggests the general reliability of rm and qm based on
the mean- $\text{WA}_{\epsilon,\gamma}$ -median inequality for unimodal distribu-
tions. Extending this rationale to other weighted L -statistics
is possible, since the U -orderliness can also be bounded with
certain assumptions, as discussed previously.

B. Robust estimations of the central moments. In 1979, Bickel
and Lehmann (13), in their final paper of the landmark series
Descriptive Statistics for Nonparametric Models, generalized a
class of estimators called measures of spread, which "do not
require the assumption of symmetry." From this, a popular
efficient scale estimator, the Rousseeuw-Croux scale estimator
(14), was derived in 1993, but the importance of tackling
the symmetry assumption has been greatly underestimated.
While they had already considered one version of the trimmed
standard deviation, which is a measures of dispersion, in
the third paper of that series (15); in the final section of
that paper (13), they explored another two versions of the
trimmed standard deviation based on pairwise differences, one
is modified here for comparison,

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

where $(X - X')_1 \leq \dots \leq (X - X')_{\binom{n}{2}}$ are the order statistics
of the pseudo-sample, $X_i - X_j$, $i < j$. They showed that,
when $\epsilon = 0$, [1] is $\sqrt{2}$ times the standard deviation. The paper
ended with, "We do not know a fortiori which of the measures
is preferable and leave these interesting questions open." To
address their open question, the nomenclature used in this
paper is introduced as follows:

Nomenclature. Given a robust estimator $\hat{\theta}$, which has an
adjustable breakdown point that can approach zero asymptoti-
cally, the name of $\hat{\theta}$ comprises two parts: the first part denotes
the type of estimator, and the second part is the name of the
population parameter θ , with which the estimator is consistent
as $\epsilon \rightarrow 0$. The abbreviation of the estimator combines the
initial letters of the first part and the population parameter. If
the estimator is symmetric, the upper asymptotic breakdown

point, ϵ (defined in Subsection ??, or ϵ_{U_k} for LU -statistic), is indicated in the subscript of the abbreviation of the estimator, with the exception of the median. For an asymmetric estimator based on quantile average, the associated γ follows ϵ .

In the previous article on semiparametric robust mean estimation, it was shown that the bias of a robust estimator with an adjustable breakdown point is often monotonic with respect to the breakdown point in a semiparametric distribution. Naturally, the estimator's name should reflect the population parameter with which it is consistent as $\epsilon \rightarrow 0$. If multiplying all pseudo-samples by a factor of $\frac{1}{\sqrt{2}}$, then [1] is the trimmed standard deviation adhering to this nomenclature, since $\psi_2(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$ is the kernel function of the unbiased estimation of the second central moment by using U -statistic. It should be preferable, not only because it is the square root of a trimmed U -statistic, which is closely related to the minimum-variance unbiased estimator (MVUE), but also because the second 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 unimodal distribution is second γ -ordered, if $\gamma \geq 0$.*

Proof. In 1954, Hodges and Lehmann established that if X and Y are independently drawn from the same unimodal distribution, $X - Y$ will be a symmetric unimodal distribution peaking at zero (16). Given the constraint in the pseudo-sample that $X_i < X_j$, $i < j$, it directly follows from Theorem 1 in (16) that the pairwise difference distribution (Ξ) generated from any unimodal distribution is always monotonic increasing with a mode at zero. The transformation of the pairwise difference distribution via squaring and multiplication by $\frac{1}{2}$ does not change the monotonicity, making the pdf to become monotonically decreasing with a mode at zero. In the previous article, it was proven that a right-skewed distribution with a monotonic decreasing pdf is always second γ -ordered for non-negative γ , which gives the desired result. \square

Previously, it was shown that any symmetric distribution with a finite second moment is ν th ordered, suggesting that ν th orderliness does not require unimodality, e.g., a symmetric bimodal distribution is also ν th ordered. An analysis of the Weibull distribution showed that unimodality does not assure orderliness. Theorem B.1 uncovers a profound relationship between unimodality and second γ -orderliness, which is sufficient for γ -trimming inequality and γ -orderliness.

In 1928, Fisher constructed k -statistics as unbiased estimators of cumulants (17). Halmos (1946) proved that a 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, unbiasedness and minimum variance (18). Hoeffding, in 1948, generalized U -statistics (19) which enable the derivation of a minimum-variance unbiased estimator from each unbiased estimator of an estimable parameter. 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 trimmed U -statistics (20). Replacing the trimmed mean in the trimmed U -statistics with LL -statistic forms the definition of LU -statistics.

In 1997, Heffernan (21) obtained an unbiased estimator of the k th central moment by using U -statistics and demonstrated that it is the minimum variance unbiased estimator for distributions with the finite first k moments. The weighted k th central moment ($2 \leq k \leq n$) is thus defined as,

$$Wkm_{\epsilon_{U_k}, \gamma, n} := WL_{\epsilon, \gamma, n} \left((\psi_k(X_{N_1}, \dots, X_{N_k}))_{N=1}^{\binom{n}{k}} \right),$$

where $\epsilon_{U_k} = 1 - (1 - \epsilon)^{\frac{1}{k}}$, X_{N_1}, \dots, X_{N_k} are the n choose k elements from the sample, $\psi_k(x_1, \dots, x_k) = \sum_{j=0}^{k-2} (-1)^j \binom{k-1}{k-j} \sum (x_{i_1}^{k-j} x_{i_2} \dots x_{i_{j+1}}) + (-1)^{k-1} (k-1) x_1 \dots x_k$, the second summation is over $i_1, \dots, i_{j+1} = 1$ to k with $i_1 \neq i_2 \neq \dots \neq i_{j+1}$ and $i_2 < i_3 < \dots < i_{j+1}$ (21). Despite the complexity, the following theorem offers an approach to infer the general structure of such kernel distributions.

Theorem B.2. *For each pair $(Q(p_1), Q(p_k))$ derived from the original distribution where $Q(p_1) < Q(p_k)$, let $Q(p_1)$ and $Q(p_k)$ be the first and last inputs to the kernel function $\psi_k(x_1, \dots, x_k)$. Define $\Delta = Q(p_1) - Q(p_k)$ and denote ξ_Δ be the component quasi-distribution which is a function of Δ , as detailed below. The k th, where $k > 2$, central moment kernel distribution, labeled Ξ_k , can be seen as a quasi-mixture distribution with an infinite number of components. Each component possesses the support $(-\binom{k}{3+(-1)^k}^{-1}(-\Delta)^k, \frac{1}{k}(-\Delta)^k)$.*

Proof. Without loss of generality, generating the distribution of the k -tuple $(Q(p_1), \dots, Q(p_k))$ under continuity, $k > 2$, $1 < \dots < k$, $p_1 < \dots < p_k$, the corresponding probability density is $f_{X, \dots, X}(Q(p_1), \dots, Q(p_k)) = k! f(Q(p_1)) \dots f(Q(p_k))$ (a result after a modification of the Jacobian density theorem). Transforming the distribution of the k -tuple by the function $\psi_k(x_1, \dots, x_k)$, denoting $\bar{\Delta} = \psi_k(Q(p_1), \dots, Q(p_k))$. The probability $f_{\bar{\Delta}}(\bar{\Delta}) = \sum_{\bar{\Delta} = \psi_k(Q(p_1), \dots, Q(p_k))} f_{X, \dots, X}(Q(p_1), \dots, Q(p_k))$ is the summation of the probabilities of all k -tuples that satisfy $\bar{\Delta} = \psi_k(Q(p_1), \dots, Q(p_k))$. The following Ξ_k is equivalent.

Ξ_k : Every pair with a difference equal to $\Delta = Q(p_1) - Q(p_k)$ can generate a quasi-distribution (but the integral is not equal to 1, so quasi) from the distribution of the k -tuple by the function ψ_k , such that $Q(p_2), \dots, Q(p_{k-1})$ exhaust all combinations under the inequality constraints, i.e., $Q(p_1) < Q(p_2) < \dots < Q(p_{k-1}) < Q(p_k)$. Combing all the quasi-distributions with the same Δ forms ξ_Δ . Ξ_k is a quasi-mixture distribution that can be expressed as an integral of ξ_Δ over all possible values of Δ , ranging from $Q(0) - Q(1)$ to 0. The probability of ξ_Δ , $f_{\xi_\Delta}(\bar{\Delta}) = f_{\Xi_k}(\bar{\Delta}|\Delta)$. \square

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. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
4. X He, WK Fung, Method of medians for lifetime data with weibull models. *Stat. medicine* **18**, 1993–2009 (1999).

- 377 5. M Menon, Estimation of the shape and scale parameters of the weibull distribution. *Techno-*
378 *metrics* **5**, 175–182 (1963).
- 379 6. SD Dubey, Some percentile estimators for weibull parameters. *Technometrics* **9**, 119–129
380 (1967).
- 381 7. NB Marks, Estimation of weibull parameters from common percentiles. *J. applied Stat.* **32**,
382 17–24 (2005).
- 383 8. K Boudt, D Caliskan, C Croux, Robust explicit estimators of weibull parameters. *Metrika* **73**,
384 187–209 (2011).
- 385 9. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under
386 partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
- 387 10. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
- 388 11. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
- 389 12. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D*
390 *(The Stat.* **33**, 391–399 (1984).
- 391 13. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models iv. spread in *Selected*
392 *Works of EL Lehmann.* (Springer), pp. 519–526 (2012).
- 393 14. PJ Rousseeuw, C Croux, Alternatives to the median absolute deviation. *J. Am. Stat. associa-*
394 *tion* **88**, 1273–1283 (1993).
- 395 15. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models. iii. dispersion in
396 *Selected works of EL Lehmann.* (Springer), pp. 499–518 (2012).
- 397 16. J Hodges, E Lehmann, Matching in paired comparisons. *The Annals Math. Stat.* **25**, 787–791
398 (1954).
- 399 17. RA Fisher, Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.*
400 **2**, 199–238 (1930).
- 401 18. PR Halmos, The theory of unbiased estimation. *The Annals Math. Stat.* **17**, 34–43 (1946).
- 402 19. W Hoeffding, A class of statistics with asymptotically normal distribution. *The Annals Math.*
403 *Stat.* **19**, 293–325 (1948).
- 404 20. RJ Serfling, Generalized l-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).
- 405 21. PM Heffernan, Unbiased estimation of central moments by using u-statistics. *J. Royal Stat.*
406 *Soc. Ser. B (Statistical Methodol.* **59**, 861–863 (1997).

DRAFT