Robust estimations from distribution structures: II. Central Moments

Tuobang Li

This manuscript was compiled on November 28, 2023

Descriptive statistics for parametric models are currently highly sensative to departures, gross errors, and/or random errors. Here, leveraging the structures of parametric distributions and their central moment kernel distributions, a class of estimators, consistent simultanously for both a semiparametric distribution and a distinct parametric distribution, is proposed. These efficient estimators are robust to both gross errors and departures from parametric assumptions, making them ideal for estimating the mean and central moments of common unimodal distributions. This article also illuminates the understanding of the common nature of probability distributions and the measures of them.

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

The potential biases of robust location estimators in estimating the population mean have been noticed for more than two centuries (1), with numerous significant attempts made to address them. In calculating a robust estimator, the procedure of identifying and downweighting extreme values inherently necessitates the formulation of distributional assumptions. Previously, it was demonstrated that, due to the presence of infinite-dimensional nuisance shape parameters, the semiparametric approach struggles to consistently address distributions with shapes more intricate than symmetry. 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 (HM) increases rapidly. This is a common issue in parametric robust statistics. For example, He and Fung (1999) constructed (4) a robust M-estimator (HFM) for the two-parameter Weibull distribution, from which the mean and central moments can be calculated. Nonetheless, it is inadequate for other parametric distributions, e.g., the gamma, Perato, lognormal, and the generalized Gaussian distributions (SI Dataset S1). Another interesting approach is based on L-estimators, such as percentile estimators. For examples of percentile estimators for the Weibull distribution, the reader is referred to the works of Menon (1963) (5), Dubey (1967) (6), Marks (2005) (7), and Boudt, Caliskan, and Croux (2011) (8). At the outset of the study of percentile estimators, it was known that they arithmetically utilize the invariant structures of parametric distributions (5, 6). An estimator is classified as an I-statistic if it asymptotically satisfies $\mathrm{I}\left(\mathrm{LE}_1, \ldots, \mathrm{LE}_l\right) = (\theta_1, \ldots, \theta_q)$ for the distribution it is consistent, where LEs are calculated with the use of LU-statistics (defined in Subsection A), I is defined using arithmetic operations and constants but may also incorporate transcendental functions and quantile functions, and θs are the population parameters it estimates. In this article, two subclasses of I-statistics are introduced, recombined I-statistics and quantile I-statistics. 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 even when the sample follows a gamma distribution, which belongs to the same larger family as the Weibull model, the generalized gamma distribution, a misassumption can still lead to substantial biases in Marks percentile estimator (MP) for the Weibull distribution (7) (SI Dataset S1).

On the other hand, while robust estimation of scale has also been intensively studied with established methods (9, 10), the development of robust measures of asymmetry and kurtosis lags behind, despite the availability of several approaches (11–15). 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 by utilizing the invariant structures of unimodal distributions, a suite of robust estimators can be constructed whose biases are typically smaller than the variances (as seen in Table 1 for $n = 5184$).

A. Robust Estimations of the Central Moments. The most popular robust scale estimator currently, the median absolute deviation, was popularized by Hampel (1974) (16), who credits the idea to Gauss in 1816 (17). In 1976, in their landmark series *Descriptive Statistics for Nonparametric Models*, Bickel and Lehmann (9) generalized a class of estimators as measures of the dispersion of a symmetric distribution around its center of symmetry. In 1979, the same series, they (10) proposed a class of estimators referred to as measures of spread, which consider the pairwise differences of a random variable, irrespective of its symmetry, throughout its distribution, rather than

Significance Statement

Bias, variance, and contamination are the three main errors in statistics. Consistent robust estimation is unattainable without distributional 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.

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

focusing on dispersion relative to a fixed point. In the final section (10), they explored a version of the trimmed standard deviation based on pairwise differences, which is modified here for comparison,

$$\left[\binom{n}{2}(1-\epsilon_{\mathbf{0}}-\gamma\epsilon_{\mathbf{0}})\right]^{-\frac{1}{2}}\left[\sum_{i=\binom{n}{2}\gamma\epsilon_{\mathbf{0}}}^{\binom{n}{2}(1-\epsilon_{\mathbf{0}})}\left(X_{i_1}-X_{i_2}\right)_i^{\,2}\right]^{\frac{1}{2}}, \quad [1]$$

where $\left(X_{i_1}-X_{i_2}\right)_1 \leq \ldots \leq \left(X_{i_1}-X_{i_2}\right)_{\binom{n}{2}}$ are the order statistics of $X_{i_1}-X_{i_2}$, $i_1<i_2$, provided that $\binom{n}{2}\gamma\epsilon_{\mathbf{0}} \in \mathbb{N}$ and $\binom{n}{2}(1-\epsilon_{\mathbf{0}}) \in \mathbb{N}$. They showed that, when $\epsilon_{\mathbf{0}}=0$, the result obtained using [1] is equal to $\sqrt{2}$ times the sample standard deviation. The paper ended with, "We do not know a fortiori which of the measures is preferable and leave these interesting questions open."

Two examples of the impacts of that series are as follows. Oja (1981, 1983) (18, 19) provided a more comprehensive and generalized examination of these concepts, and integrated the measures of location, dispersion, and spread as proposed by Bickel and Lehmann (9, 10, 20), along with van Zwet's convex transformation order of skewness and kurtosis (1964) (21) for univariate and multivariate distributions, resulting a greater degree of generality and a broader perspective on these statistical constructs. Rousseeuw and Croux proposed a popular efficient scale estimator based on separate medians of pairwise differences taken over i_1 and i_2 (22) in 1993. However the importance of tackling the symmetry assumption has been greatly underestimated, as will be discussed later.

To address their open question (10), the nomenclature used in this paper is introduced as follows:

Nomenclature. Given a robust estimator, $\hat{\theta}$, which has an adjustable breakdown point, ϵ, that can approach zero asymptotically, the name of $\hat{\theta}$ comprises two parts: the first part denotes the type of estimator, and the second part represents the population parameter θ, such that $\hat{\theta} \to \theta$ as $\epsilon \to 0$. The abbreviation of the estimator combines the initial letters of the first part and the second part. If the estimator is symmetric, the upper asymptotic breakdown point, ϵ, 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 REDS I, 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 that it approaches as $\epsilon \to 0$. 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 (23). This definition 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 A.1. *The second central moment kernel distribution generated from any unimodal distribution is second γ-ordered, provided that $\gamma \geq 0$.*

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 (24). Given the constraint in the pairwise differences that $X_{i_1}<X_{i_2}$, $i_1<i_2$, it directly follows from Theorem 1 in (24) that the pairwise difference distribution (Ξ_{Δ}) generated from any unimodal distribution is always monotonic increasing with a mode at zero. Since $X-X'$ is a negative variable that is monotonically increasing, applying the squaring transformation, the relationship between the original variable $X-X'$ and its squared counterpart $(X-X')^2$ can be represented as follows: $X-X'<Y-Y' \implies (X-X')^2>(Y-Y')^2$. In other words, as the negative values of $X-X'$ become larger in magnitude (more negative), their squared values $(X-X')^2$ become larger as well, but in a monotonically decreasing manner with a mode at zero. Further multiplication by $\frac{1}{2}$ also does not change the monotonicity and mode, since the mode is zero. Therefore, the transformed pdf becomes monotonically decreasing with a mode at zero. In REDS I, it was proven that a right-skewed distribution with a monotonic decreasing pdf is always second γ-ordered, which gives the desired result. ☐

In REDS I, it was shown that any symmetric distribution is νth U-ordered, suggesting that νth U-orderliness does not require unimodality, e.g., a symmetric bimodal distribution is also νth U-ordered. In the SI Text of REDS I, an analysis of the Weibull distribution showed that unimodality does not assure orderliness. Theorem A.1 uncovers a profound relationship between unimodality, monotonicity, and second γ-orderliness, which is sufficient for γ-trimming inequality and γ-orderliness.

In 1928, Fisher constructed $\mathbf{k}$-statistics as unbiased estimators of cumulants (25). Halmos (1946) proved that a functional θ admits an unbiased estimator if and only if it is a regular statistical functional of degree $\mathbf{k}$ and showed a relation of symmetry, unbiasness and minimum variance (26). Hoeffding, in 1948, generalized U-statistics (27) 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 (28). Given a kernel function $h_{\mathbf{k}}$ which is a symmetric function of $\mathbf{k}$ variables, the LU-statistic is defined as:

$$LU_{h_{\mathbf{k}},\mathbf{k},k,\epsilon,\gamma,n} := LL_{k,\epsilon_{\mathbf{0}},\gamma,n}\left(\operatorname{sort}\left(\left(h_{\mathbf{k}}\left(X_{N_1},\ldots,X_{N_{\mathbf{k}}}\right)\right)_{N=1}^{\binom{n}{\mathbf{k}}}\right)\right),$$

where $\epsilon = 1-(1-\epsilon_{\mathbf{0}})^{\frac{1}{\mathbf{k}}}$ (proven in Subsection B), $X_{N_1},\ldots,X_{N_{\mathbf{k}}}$ are the n choose $\mathbf{k}$ elements from the sample, $LL_{k,\epsilon_{\mathbf{0}},\gamma,n}(Y)$ denotes the LL-statistic with the sorted sequence $\operatorname{sort}\left(\left(h_{\mathbf{k}}\left(X_{N_1},\ldots,X_{N_{\mathbf{k}}}\right)\right)_{N=1}^{\binom{n}{\mathbf{k}}}\right)$ serving as an input. In the context of Serfling's work, the term 'trimmed U-statistic' is used when $LL_{k,\epsilon_{\mathbf{0}},\gamma,n}$ is $\mathrm{TM}_{\epsilon_{\mathbf{0}},\gamma,n}$ (28).

In 1997, Heffernan (23) obtained an unbiased estimator of the $\mathbf{k}$th central moment by using U-statistics and demonstrated that it is the minimum variance unbiased estimator for distributions with the finite first $\mathbf{k}$ moments. The weighted H-L $\mathbf{k}$th central moment ($2 \leq \mathbf{k} \leq n$) is thus defined as,

$$\mathrm{WHL}\mathbf{k}m_{k,\epsilon,\gamma,n} := LU_{h_{\mathbf{k}}=\psi_{\mathbf{k}},\mathbf{k},k,\epsilon,\gamma,n},$$

where $\mathrm{WHLM}_{k,\epsilon_{\mathbf{0}},\gamma,n}$ is used as the $LL_{k,\epsilon_{\mathbf{0}},\gamma,n}$ in LU, $\psi_{\mathbf{k}}\left(x_1,\ldots,x_{\mathbf{k}}\right)=\sum_{j=0}^{\mathbf{k}-2}\left(-1\right)^j\left(\frac{1}{\mathbf{k}-j}\right)\sum\left(x_{i_1}^{\mathbf{k}-j}x_{i_2}\ldots x_{i_{j+1}}\right)+\left(-1\right)^{\mathbf{k}-1}\left(\mathbf{k}-1\right)x_1\ldots x_{\mathbf{k}}$, the second summation is over $i_1,\ldots,i_{j+1}=1$ to $\mathbf{k}$ with $i_1\neq i_2\neq\ldots\neq i_{j+1}$ and $i_2<i_3<\ldots<i_{j+1}$ (23). Despite the complexity, the following theorem offers an approach to infer the general structure of such kernel distributions.

Theorem A.2. *Define a set T comprising all pairs $(\psi_{\mathbf{k}}(\mathbf{v}),f_{X,\ldots,X}(\mathbf{v}))$ such that $\psi_{\mathbf{k}}(\mathbf{v})=\psi_{\mathbf{k}}\left(Q(p_1),\ldots,Q(p_{\mathbf{k}})\right)$ with $Q(p_1)<\ldots<Q(p_{\mathbf{k}})$ and $f_{X,\ldots,X}(\mathbf{v})=\mathbf{k}!f(Q(p_1))\ldots f(Q(p_{\mathbf{k}}))$ is the probability density of the $\mathbf{k}$-tuple, $\mathbf{v}=(Q(p_1),\ldots,Q(p_{\mathbf{k}}))$ (a formula drawn after a modification of the Jacobian density theorem). T_Δ is a subset of T, consisting all those pairs for which the corresponding $\mathbf{k}$-tuples satisfy that $Q(p_1)-Q(p_{\mathbf{k}})=\Delta$. The component quasi-distribution, denoted by ξ_Δ, has a quasi-pdf $f_{\xi_\Delta}(\bar{\Delta})=\sum_{\substack{(\psi_{\mathbf{k}}(\mathbf{v}),f_{X,\ldots,X}(\mathbf{v}))\in T_\Delta\\ \bar{\Delta}=\psi_{\mathbf{k}}(\mathbf{v})}}f_{X,\ldots,X}(\mathbf{v})$, i.e., sum over all $f_{X,\ldots,X}(\mathbf{v})$ such that the pair $(\psi_{\mathbf{k}}(\mathbf{v}),f_{X,\ldots,X}(\mathbf{v}))$ is in the set T_Δ and the first element of the pair, $\psi_{\mathbf{k}}(\mathbf{v})$, is equal to $\bar{\Delta}$. The $\mathbf{k}$th, where $\mathbf{k}>2$, central moment kernel distribution, labeled $\Xi_{\mathbf{k}}$, can be seen as a quasi-mixture distribution comprising an infinite number of component quasi-distributions, ξ_Δs, each corresponding to a different value of Δ, which ranges from $Q(0)-Q(1)$ to 0. Each component quasi-distribution has a support of $\left(-\left(\frac{\mathbf{k}}{\frac{3+(-1)^{\mathbf{k}}}{2}}\right)^{-1}(-\Delta)^{\mathbf{k}},\frac{1}{\mathbf{k}}(-\Delta)^{\mathbf{k}}\right)$.*

Proof. The support of ξ_Δ is the extrema of the function $\psi_{\mathbf{k}}\left(Q(p_1),\cdots,Q(p_{\mathbf{k}})\right)$ subjected to the constraints, $Q(p_1)<\cdots<Q(p_{\mathbf{k}})$ and $\Delta=Q(p_1)-Q(p_{\mathbf{k}})$. Using the Lagrange multiplier, the only critical point can be determined at $Q(p_1)=\cdots=Q(p_{\mathbf{k}})=0$, where $\psi_{\mathbf{k}}=0$. Other candidates are within the boundaries, i.e., $\psi_{\mathbf{k}}\left(x_1=Q(p_1),x_2=Q(p_{\mathbf{k}}),\cdots,x_{\mathbf{k}}=Q(p_{\mathbf{k}})\right)$, $\cdots$, $\psi_{\mathbf{k}}\left(x_1=Q(p_1),\cdots,x_i=Q(p_1),x_{i+1}=Q(p_{\mathbf{k}}),\cdots,x_{\mathbf{k}}=Q(p_{\mathbf{k}})\right)$, $\cdots$, $\psi_{\mathbf{k}}\left(x_1=Q(p_1),\cdots,x_{\mathbf{k}-1}=Q(p_1),x_{\mathbf{k}}=Q(p_{\mathbf{k}})\right)$. $\psi_{\mathbf{k}}\left(x_1=Q(p_1),\cdots,x_i=Q(p_1),x_{i+1}=Q(p_{\mathbf{k}}),\cdots,x_{\mathbf{k}}=Q(p_{\mathbf{k}})\right)$ can be divided into $\mathbf{k}$ groups. The gth group has the common factor $\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}$, if $1\leq g\leq\mathbf{k}-1$ and the final $\mathbf{k}$th group is the term $\left(-1\right)^{\mathbf{k}-1}\left(\mathbf{k}-1\right)Q(p_1)^iQ(p_{\mathbf{k}})^{\mathbf{k}-i}$. If $\frac{\mathbf{k}+1-i}{2}\leq j\leq\frac{\mathbf{k}-1}{2}$ and $j+1\leq g\leq\mathbf{k}-j$, the gth group has $i\binom{i-1}{g-j-1}\binom{\mathbf{k}-i}{j}$ terms having the form $\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}Q(p_1)^{\mathbf{k}-j}Q(p_{\mathbf{k}})^j$. If $\frac{\mathbf{k}+1-i}{2}\leq j\leq\frac{\mathbf{k}-1}{2}$ and $\mathbf{k}-j+1\leq g\leq i+j$, the gth group has $i\binom{i-1}{g-j-1}\binom{\mathbf{k}-i}{j}+\left(\mathbf{k}-i\right)\binom{\mathbf{k}-i-1}{j-\mathbf{k}+g-1}\binom{i}{\mathbf{k}-j}$ terms having the form $\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}Q(p_1)^{\mathbf{k}-j}Q(p_{\mathbf{k}})^j$. If $0\leq j<\frac{\mathbf{k}+1-i}{2}$ and $j+1\leq g\leq i+j$, the gth group has $i\binom{i-1}{g-j-1}\binom{\mathbf{k}-i}{j}$ terms having the form $\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}Q(p_1)^{\mathbf{k}-j}Q(p_{\mathbf{k}})^j$. If $\frac{\mathbf{k}}{2}\leq j\leq\mathbf{k}$ and $\mathbf{k}-j+1\leq g\leq j$, the gth group has $\left(\mathbf{k}-i\right)\binom{\mathbf{k}-i-1}{j-\mathbf{k}+g-1}\binom{i}{\mathbf{k}-j}$ terms having the form $\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}Q(p_1)^{\mathbf{k}-j}Q(p_{\mathbf{k}})^j$. If $\frac{\mathbf{k}}{2}\leq j\leq\mathbf{k}$ and $j+1\leq g\leq j+i<\mathbf{k}$, the gth group has $i\binom{i-1}{g-j-1}\binom{\mathbf{k}-i}{j}+\left(\mathbf{k}-i\right)\binom{\mathbf{k}-i-1}{j-\mathbf{k}+g-1}\binom{i}{\mathbf{k}-j}$ terms having the form $\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}Q(p_1)^{\mathbf{k}-j}Q(p_{\mathbf{k}})^j$. So, if $i+j=\mathbf{k}$, $\frac{\mathbf{k}}{2}\leq j\leq\mathbf{k}$, $0\leq i\leq\frac{\mathbf{k}}{2}$, the summed coefficient of $Q(p_1)^iQ(p_{\mathbf{k}})^{\mathbf{k}-i}$ is $\left(-1\right)^{\mathbf{k}-1}\left(\mathbf{k}-1\right)+\sum_{g=i+1}^{\mathbf{k}-1}\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}\left(\mathbf{k}-i\right)\binom{\mathbf{k}-i-1}{g-i-1}+\sum_{g=\mathbf{k}-i+1}^{\mathbf{k}-1}\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}i\binom{i-1}{g-\mathbf{k}+i-1}=\left(-1\right)^{\mathbf{k}-1}\left(\mathbf{k}-1\right)+\left(-1\right)^{\mathbf{k}+1}+\left(\mathbf{k}-i\right)\left(-1\right)^{\mathbf{k}}+\left(-1\right)^{\mathbf{k}}\left(i-1\right)=\left(-1\right)^{\mathbf{k}+1}$. The summation identities are $\sum_{g=i+1}^{\mathbf{k}-1}\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}\left(\mathbf{k}-i\right)\binom{\mathbf{k}-i-1}{g-i-1}=\left(\mathbf{k}-i\right)\int_0^1\sum_{g=i+1}^{\mathbf{k}-1}\left(-1\right)^{g+1}\binom{\mathbf{k}-i-1}{g-i-1}t^{\mathbf{k}-g}dt=\left(\mathbf{k}-i\right)\int_0^1\left(\left(-1\right)^i\left(t-1\right)^{\mathbf{k}-i-1}-\left(-1\right)^{\mathbf{k}+1}\right)dt=\left(\mathbf{k}-i\right)\left(\frac{(-1)^{\mathbf{k}}}{i-\mathbf{k}}+\left(-1\right)^{\mathbf{k}}\right)=\left(-1\right)^{\mathbf{k}+1}+\left(\mathbf{k}-i\right)\left(-1\right)^{\mathbf{k}}$ and $\sum_{g=\mathbf{k}-i+1}^{\mathbf{k}-1}\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}i\binom{i-1}{g-\mathbf{k}+i-1}=\int_0^1\sum_{g=\mathbf{k}-i+1}^{\mathbf{k}-1}\left(-1\right)^{g+1}i\binom{i-1}{g-\mathbf{k}+i-1}t^{\mathbf{k}-g}dt=\int_0^1\left(i\left(-1\right)^{\mathbf{k}-i}\left(t-1\right)^{i-1}-i\left(-1\right)^{\mathbf{k}+1}\right)dt=\left(-1\right)^{\mathbf{k}}\left(i-1\right)$. If $0\leq j<\frac{\mathbf{k}+1-i}{2}$ and $i=\mathbf{k}$, $\psi_{\mathbf{k}}=0$. If $\frac{\mathbf{k}+1-i}{2}\leq j\leq\frac{\mathbf{k}-1}{2}$ and $\frac{\mathbf{k}+1}{2}\leq i\leq\mathbf{k}-1$, the summed coefficient of $Q(p_1)^iQ(p_{\mathbf{k}})^{\mathbf{k}-i}$ is $\left(-1\right)^{\mathbf{k}-1}\left(\mathbf{k}-1\right)+\sum_{g=\mathbf{k}-i+1}^{\mathbf{k}-1}\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}i\binom{i-1}{g-\mathbf{k}+i-1}+\sum_{g=i+1}^{\mathbf{k}-1}\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}\left(\mathbf{k}-i\right)\binom{\mathbf{k}-i-1}{g-i-1}$, the same as above. If $i+j<\mathbf{k}$, since $\binom{i}{\mathbf{k}-j}=0$, the related terms can be ignored, so, using the binomial theorem and beta function, the summed coefficient of $Q(p_1)^{k-j}Q(p_{\mathbf{k}})^j$ is $\sum_{g=j+1}^{i+j}\left(-1\right)^{g+1}\frac{1}{\mathbf{k}-g+1}i\binom{i-1}{g-j-1}\binom{\mathbf{k}-i}{j}=i\binom{\mathbf{k}-i}{j}\int_0^1\sum_{g=j+1}^{i+j}\left(-1\right)^{g+1}\binom{i-1}{g-j-1}t^{\mathbf{k}-g}dt=\binom{\mathbf{k}-i}{j}i\int_0^1\left(\left(-1\right)^jt^{\mathbf{k}-j-1}\left(\frac{t}{t-1}\right)^{1-i}\right)dt=\binom{\mathbf{k}-i}{j}i\frac{(-1)^{j+i+1}\Gamma(i)\Gamma(\mathbf{k}-j-i+1)}{\Gamma(\mathbf{k}-j+1)}=\frac{(-1)^{j+i+1}i!(\mathbf{k}-j-i)!(\mathbf{k}-i)!}{(\mathbf{k}-j)!j!(\mathbf{k}-j-i)!}=\left(-1\right)^{j+i+1}\frac{i!(\mathbf{k}-i)!}{\mathbf{k}!}\frac{\mathbf{k}!}{(\mathbf{k}-j)!j!}=\binom{\mathbf{k}}{i}^{-1}\left(-1\right)^{1+i}\binom{\mathbf{k}}{j}\left(-1\right)^j$.

According to the binomial theorem, the coefficient of $Q(p_1)^iQ(p_{\mathbf{k}})^{\mathbf{k}-i}$ in $\binom{\mathbf{k}}{i}^{-1}\left(-1\right)^{1+i}\left(Q(p_1)-Q(p_{\mathbf{k}})\right)^{\mathbf{k}}$ is $\binom{\mathbf{k}}{i}^{-1}\left(-1\right)^{1+i}\binom{\mathbf{k}}{i}\left(-1\right)^{\mathbf{k}-i}=\left(-1\right)^{\mathbf{k}+1}$, same as the above summed coefficient of $Q(p_1)^iQ(p_{\mathbf{k}})^{\mathbf{k}-i}$, if $i+j=\mathbf{k}$. If $i+j<k$, the coefficient of $Q(p_1)^{\mathbf{k}-j}Q(p_{\mathbf{k}})^j$ is $\binom{\mathbf{k}}{i}^{-1}\left(-1\right)^{1+i}\binom{\mathbf{k}}{j}\left(-1\right)^j$, same as the corresponding summed coefficient of $Q(p_1)^{\mathbf{k}-j}Q(p_{\mathbf{k}})^j$. Therefore, $\psi_{\mathbf{k}}\left(x_1=Q(p_1),\ldots,x_i=Q(p_1),x_{i+1}=Q(p_{\mathbf{k}}),\ldots,x_{\mathbf{k}}=Q(p_{\mathbf{k}})\right)=\binom{\mathbf{k}}{i}^{-1}\left(-1\right)^{1+i}\left(Q(p_1)-Q(p_{\mathbf{k}})\right)^{\mathbf{k}}$, the maximum and minimum of $\psi_{\mathbf{k}}$ follow directly from the properties of the binomial coefficient.

□

The component quasi-distribution, ξ_Δ, is closely related to Ξ_Δ, which is the pairwise difference distribution, since $\sum_{\bar{\Delta}=-\left(\frac{\mathbf{k}}{\frac{3+(-1)^{\mathbf{k}}}{2}}\right)^{-1}(-\Delta)^{\mathbf{k}}}^{\frac{1}{\mathbf{k}}(-\Delta)^{\mathbf{k}}}f_{\xi_\Delta}(\bar{\Delta})=f_{\Xi_\Delta}(\Delta)$. Recall that Theorem A.1 established that $f_{\Xi_\Delta}(\Delta)$ is monotonic increasing with a mode at zero if the original distribution is unimodal, $f_{\Xi_{-\Delta}}(-\Delta)$ is thus monotonic decreasing with a mode at zero. In general, if assuming the shape of ξ_Δ is uniform, $\Xi_{\mathbf{k}}$ is monotonic left and right around zero. The median of $\Xi_{\mathbf{k}}$ 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 $-\Delta$ 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 $-\Delta$ 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 $\Xi_{\mathbf{k}}$ will only be slightly shifted from zero. Denote the median of $\Xi_{\mathbf{k}}$ as $m\mathbf{k}m$, for the five parametric distributions here, $|m\mathbf{k}m|$s are all $\leq 0.1\sigma$ for Ξ_3 and Ξ_4, where σ is the standard deviation of $\Xi_{\mathbf{k}}$ (SI Dataset S1). Assuming $m\mathbf{k}m=0$, 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

$\frac{\frac{1}{2}}{\binom{\mathbf{k}}{2}^{-1}(Q(0)-Q(1))^{\mathbf{k}}} > \frac{\frac{1}{2}}{\frac{1}{\mathbf{k}}(Q(0)-Q(1))^{\mathbf{k}}}$. 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. In all, the monotonic decreasing of the negative pairwise difference distribution guides the general shape of the $\mathbf{k}$th central moment kernel distribution, $\mathbf{k} > 2$, 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 νth γ-ordered and all of its central moment kernel distributions are also νth γ-ordered, it is called completely νth γ-ordered. Although strict complete νth orderliness is difficult to prove, even if the inequality may be violated in a small range, as discussed in Subsection **??**, the mean-SWA$_\epsilon$-median inequality remains valid, in most cases, for the central moment kernel distribution.

The next theorem shows an interesting relation between congruence and the central moment kernel distribution.

Theorem A.3. *The second central moment kernal distribution derived from a continuous location-scale unimodal distribution is always γ-congruent.*

Proof. Theorem **??** shows that the central moment kernel distribution generated from a location-scale distribution is also a location-scale distribution. Theorem A.1 shows that it is positively definite. Implementing Theorem 12 in REDS 1 yields the desired result. □

Although some parametric distributions are not congruent, as shown in REDS 1. In REDS 1, Theorem 12 establishes that γ-congruence always holds for a positive definite location-scale family distribution and thus for the second central moment kernel distribution generated from a location-scale unimodal distribution as shown in Theorem A.3. Theorem A.2 demonstrates that all central moment kernel distributions are unimodal-like with mode and median close to zero, as long as they are generated from unimodal distributions. Assuming finite moments and constant $Q(0) - Q(1)$, increasing the mean of a distribution will result in a generally more heavy-tailed distribution, i.e., the probability density of the values close to $Q(1)$ increases, since the total probability density is 1. In the case of the $\mathbf{k}$th central moment kernel distribution, $\mathbf{k} > 2$, while the total probability density on either side of zero remains generally constant as the median is generally close to zero and much less impacted by increasing the mean, the probability density of the values close to zero decreases as the mean increases. This transformation will increase nearly all symmetric weighted averages, in the general sense. Therefore, except for the median, which is assumed to be zero, nearly all symmetric weighted averages for all central moment kernel distributions derived from unimodal distributions should change in the same direction when the parameters change.

To avoid confusion, it should be noted that the robust location estimations of the kernel distributions discussed in this paper differ from the approach taken by Joly and Lugosi (2016) (29), which is computing the median of all U-statistics from different disjoint blocks. Compared to bootstrap median U-statistics, this approach can produce two additional kinds of finite sample bias, one arises from the limited numbers of blocks, another is due to the size of the U-statistics (consider the mean of all U-statistics from different disjoint blocks, it is definitely not identical to the original U-statistic, except when the kernel is the Hodges-Lehmann kernel). Laforgue, Clemencon, and Bertail (2019)'s median of randomized U-statistics (30) is more sophisticated and can overcome the limitation of the number of blocks, but the second kind of bias remains unsolved.

B. Robustness. The measure of robustness to gross errors used in this series is the breakdown point proposed by Hampel (31) in 1968. In REDS I, it has shown that the median of means (MoM) is asymptotically equivalent to the median Hodge-Lehmann mean. Therefore it is also biased for any asymmetric distribution. However, the concentration bound of MoM depends on $\sqrt{\frac{1}{n}}$ (32), it is quite natural to deduce that it is a consistent robust estimator. The concept, sample-dependent breakdown point, is defined to avoid ambiguity.

Definition B.1 (Sample-dependent breakdown point). The breakdown point of an estimator $\hat{\theta}$ is called sample-dependent if and only if the upper and lower asymptotic breakdown points, which are the upper and lower breakdown points when $n \to \infty$, are zero and the empirical influence function of $\hat{\theta}$ is bounded. For a full formal definition of the empirical influence function, the reader is referred to Devlin, Gnanadesikan and Kettenring (1975)'s paper (33).

Bear in mind that it differs from the "infinitesimal robustness" defined by Hampel, which is related to whether the asymptotic influence function is bounded (16, 34, 35). The proof of the consistency of MoM assumes that it is an estimator with a sample-dependent breakdown point since its breakdown point is $\frac{b}{2n}$, where b is the number of blocks, then $\lim_{n\to\infty}\left(\frac{b}{2n}\right) = 0$, if b is a constant and any changes in any one of the points of the sample cannot break down this estimator.

For the LU-statistics, the asymptotic upper breakdown points are suggested by the following theorem, which extends the method in Donoho and Huber (1983)'s proof of the breakdown point of the Hodges-Lehmann estimator (36). The proof is given in the SI Text.

Theorem B.1. *Given a U-statistic associated with a symmetric kernel of degree $\mathbf{k}$. Then, assuming that as $n \to \infty$, $\mathbf{k}$ is a constant, the upper breakdown point of the LU-statistic is $1 - (1-\epsilon_{\mathbf{0}})^{\frac{1}{\mathbf{k}}}$, where $\epsilon_{\mathbf{0}}$ is the upper breakdown point of the corresponding LL-statistic.*

Remark. If $\mathbf{k} = 1$, $1 - (1-\epsilon_{\mathbf{0}})^{\frac{1}{\mathbf{k}}} = \epsilon_{\mathbf{0}}$, so this formula also holds for the LL-statistic itself. Here, to ensure the breakdown points of all four moments are the same, $\frac{1}{24}$, since $\epsilon_{\mathbf{0}} = 1 - (1-\epsilon)^{\mathbf{k}}$, the breakdown points of all LU-statistics for the second, third, and fourth central moment estimations are adjusted as $\epsilon_{\mathbf{0}} = \frac{47}{576}, \frac{1657}{13824}, \frac{51935}{331776}$, respectively.

Every statistic is based on certain assumptions. For instance, the sample mean assumes that the second moment

Table 1. Evaluation of invariant moments for five common unimodal distributions in comparison with current popular methods

Errors	$\bar{x}$	TM	H-L	SM	HM	WM	SQM	BM	MoM	MoRM	mHLM	$rm_{exp,\mathrm{BM}}$	$qm_{exp,\mathrm{BM}}$
WASAB	0.000	0.107	0.088	0.078	0.078	0.066	0.048	0.048	0.034	0.035	0.034	0.002	0.003
WRMSE	0.014	0.111	0.092	0.083	0.083	0.070	0.053	0.053	0.041	0.041	0.038	0.017	0.018
WASB$_{n=5184}$	0.000	0.108	0.089	0.078	0.079	0.066	0.048	0.048	0.034	0.036	0.033	0.002	0.003
WSE $\vee$ WSSE	0.014	0.014	0.014	0.015	0.014	0.014	0.014	0.015	0.017	0.014	0.014	0.017	0.017

Errors		HFM$_{\mu}$	MP$_{\mu}$	rm	qm	im	var	var_{bs}	Tsd^2	HFM$_{\mu_2}$	MP$_{\mu_2}$	$rvar$	$qvar$	$ivar$
WASAB		0.037	0.043	0.001	0.002	0.001	0.000	0.000	0.200	0.027	0.042	0.005	0.018	0.003
WRMSE		0.049	0.055	0.015	0.015	0.014	0.017	0.017	0.198	0.042	0.062	0.019	0.026	0.019
WASB$_{n=5184}$		0.038	0.043	0.001	0.002	0.001	0.000	0.001	0.198	0.027	0.043	0.005	0.018	0.003
WSE $\vee$ WSSE		0.018	0.021	0.015	0.015	0.014	0.017	0.017	0.015	0.024	0.032	0.018	0.017	0.018

Errors	tm	tm_{bs}	HFM$_{\mu_3}$	MP$_{\mu_3}$	rtm	qtm	itm	fm	fm_{bs}	HFM$_{\mu_4}$	MP$_{\mu_4}$	rfm	qfm	ifm
WASAB	0.000	0.000	0.052	0.059	0.006	0.083	0.034	0.000	0.000	0.037	0.046	0.024	0.038	0.011
WRMSE	0.019	0.018	0.063	0.074	0.018	0.083	0.044	0.026	0.023	0.049	0.062	0.037	0.043	0.029
WASB$_{n=5184}$	0.001	0.003	0.052	0.059	0.007	0.082	0.038	0.001	0.009	0.037	0.047	0.024	0.036	0.013
WSE $\vee$ WSSE	0.019	0.018	0.021	0.091	0.015	0.012	0.017	0.024	0.021	0.020	0.027	0.021	0.020	0.022

The first table presents the use of the exponential distribution as the consistent distribution for five common unimodal distributions: Weibull, gamma, Pareto, lognormal, and generalized Gaussian distributions. Popular robust mean estimators discussed in REDS 1 were used as comparisons. The breakdown points of mean estimators in the first table, besides H-L estimator and Huber M-estimator, are all $\frac{1}{8}$. The second and third tables present the use of the Weibull distribution as the consistent distribution not plus/plus using the lognormal distribution for the odd ordinal moments optimization and the generalized Gaussian distribution for the even ordinal moments optimization. SQM is the robust mean estimator used in recombined/quantile moments. Unbiased sample central moments (var, tm, fm), U-central moments with quasi-bootstrap (var_{bs}, tm_{bs}, fm_{bs}), and other estimators were used as comparisons. The generalized Gaussian distribution was excluded for He and Fung M-Estimator and Marks percentile estimator, since the logarithmic function does not produce results for negative inputs. The breakdown points of estimators in the second and third table, besides M-estimators and percentile estimator, are all $\frac{1}{24}$. The tables include the average standardized asymptotic bias (ASAB, as $n \to \infty$), root mean square error (RMSE, at $n = 5184$), average standardized bias (ASB, at $n = 5184$) and variance (SE $\vee$ SSE, at $n = 5184$) of these estimators, all reported in the units of the standard deviations of the distribution or corresponding kernel distributions. W means that the results were weighted by the number of Google Scholar search results on May 30, 2022 (including synonyms). The calibrations of d values and the computations of ASAB, ASB, and SSE were described in Subsection **??**, B and SI Methods. Detailed results and related codes are available in SI Dataset S1 and GitHub.

of the underlying distribution is finite. If this assumption is violated, the variance of the sample mean becomes infinitely large, even if the population mean is finite. As a result, the sample mean not only has zero robustness to gross errors, but also has zero robustness to departures. To meaningfully compare the performance of estimators under departures from assumptions, it is necessary to impose constraints on these departures. Bound analysis (1) is the first approach to study the robustness to departures, i.e., although all estimators can be biased under departures from the corresponding assumptions, but their standardized maximum deviations can differ substantially (32, 37–41). In REDS I, it is shown that another way to qualitatively compare the estimators' robustness to departures from the symmetry assumption is constructing and comparing corresponding semiparametric models. While such comparison is limited to a semiparametric model and is not universal, it is still valid for a wide range of parametric distributions. Bound analysis is a more universal approach since they can be deduced by just assuming regularity conditions (32, 37–39, 41). However, bounds are often hard to deduce for complex estimators. Also, sometimes there are discrepancies between maximum bias and average bias. Since the estimators proposed here are all consistent under certain assumptions, measuring their biases is also a convenient way of measuring the robustness to departures. Average standardized asymptotic bias is thus defined as follows.

Definition B.2 (Average standardized asymptotic bias). For a single-parameter distribution, the average standardized asymptotic bias (ASAB) is given by $\frac{|\hat{\theta}-\theta|}{\sigma}$, where $\hat{\theta}$ represents the estimation of θ, and σ denotes the standard deviation of the kernel distribution associated with the LU-statistic. If the estimator $\hat{\theta}$ is not classified as an RI-statistic, QI-statistic, or LU-statistic, the corresponding U-statistic, which measures the same attribute of the distribution, is utilized to determine the value of σ. For a two-parameter distribution, the first step is setting the lower bound of the kurtosis range of interest $\tilde{\mu}_{4_l}$, the spacing δ, and the bin count C. Then, the average standardized asymptotic bias is defined as

$$\mathrm{ASAB}_{\hat{\theta}} := \frac{1}{C} \sum_{\substack{\delta+\tilde{\mu}_{4_l} \leq \tilde{\mu}_4 \leq C\delta+\tilde{\mu}_{4_l} \\ \tilde{\mu}_4 \text{ is a multiple of } \delta}} E_{\hat{\theta}|\tilde{\mu}_4}\left[\frac{\left|\hat{\theta}-\theta\right|}{\sigma}\right]$$

where $\tilde{\mu}_4$ is the kurtosis specifying the two-parameter distribution, $E_{\hat{\theta}|\tilde{\mu}_4}$ denotes the expected value given fixed $\tilde{\mu}_4$.

Standardization plays a crucial role in comparing the performance of estimators across different distributions. Currently, several options are available, such as using the root mean square deviation from the mode (as in Gauss (1)), the mean absolute deviation, or the standard deviation. The standard deviation is preferred due to its central role in standard error estimation. In Table 1, $\delta = 0.1$, $C = 70$. For the Weibull, gamma, lognormal and generalized Gaussian distributions, $\tilde{\mu}_{4_l} = 3$ (there are two shape parameter solutions for the Weibull distribution, the lower one is used here). For the Pareto distribution, $\tilde{\mu}_{4_l} = 9$. To provide a more practical and straightforward illustration, all results from five distributions are further weighted by the number of Google Scholar search results. Within the range of kurtosis setting, nearly all WLs and WHL$\mathbf{k}m$s proposed here reach or at least come close to their maximum biases (SI Dataset S1). The pseudo-maximum bias is thus defined as the maximum value of the biases within the range of kurtosis setting for all five unimodal distributions. In most cases, the pseudo-maximum biases of invariant moments occur in lognormal or generalized Gaussian distributions (SI Dataset S1), since besides unimodality, the Weibull distribution differs entirely from them. Interestingly, the asymptotic biases of $\mathrm{TM}_{\epsilon=\frac{1}{8}}$ and $\mathrm{WM}_{\epsilon=\frac{1}{8}}$, after averaging and weighting, are 0.107σ and 0.066σ, respectively, in line with the sharp bias bounds of $\mathrm{TM}_{2,14:15}$ and $\mathrm{WM}_{2,14:15}$ (a different subscript is used to indicate a sample size of 15, with the removal of the first and last order statistics), 0.173σ and 0.126σ, for any distributions with finite moments (37, 38).

Discussion

Moments, including raw moments, central moments, and standardized moments, are the most common parameters that describe probability distributions. Central moments are preferred over raw moments because they are invariant to translation. In 1947, Hsu and Robbins proved that the arithmetic mean converges completely to the population mean provided the second moment is finite (42). The strong law of large numbers (proven by Kolmogorov in 1933) (43) implies that the $\mathbf{k}$th sample central moment is asymptotically unbiased. Recently, fascinating statistical phenomena regarding Taylor's law for distributions with infinite moments have been discovered by Drton and Xiao (2016) (44), Pillai and Meng (2016) (45), Cohen, Davis, and Samorodnitsky (2020) (46), and Brown, Cohen, Tang, and Yam (2021) (47). Lindquist and Rachev (2021) raised a critical question in their inspiring comment to Brown et al's paper (47): "What are the proper measures for the location, spread, asymmetry, and dependence (association) for random samples with infinite mean?" (48). From a different perspective, this question closely aligns with the essence of Bickel and Lehmann's open question in 1979 (10). They suggested using median, interquartile range, and medcouple (49) as the robust versions of the first three moments. While answering this question is not the focus of this paper, it is almost certain that the estimators proposed in this series will have a place. Since the efficiency of an L-statistic to the sample mean is generally monotonic with respect to the breakdown point (20), and the estimation of central moments can be transformed into the location estimation of the central moment kernel distribution, similar monotonic relations can be expected. In the case of a distribution with an infinite mean, non-robust estimators will not converge and will not provide valid estimates since their variances will be infinitely large. Therefore, the desired measures should be as robust as possible. Clearly now, if one wants to preserve the original relationship between each moment while ensuring maximum robustness, the natural choices are median, median variance, and median skewness. Similar to the robust version of L-moment (50) being trimmed L-moment (15), mean and central moments now also have their standard most robust version based on the complete congruence of the underlying distribution.

More generally, statistics, encompassing the collection, analysis, interpretation, and presentation of data, has evolved over time, with various approaches emerging to meet challenges in practice. Among these approaches, the use of probability models and measures of random variables for data analysis is often considered the core of statistics. While the early development of statistics was focused on parametric methods, there were two main approaches to point estimation. The Gauss–Markov theorem (1, 51) states the principle of minimum variance unbiased estimation which was further enriched by Neyman (1934) (52), Rao (1945) (53), Blackwell (1947) (54), and Lehmann and Scheffé (1950, 1955) (55, 56). Maximum likelihood was first introduced by Fisher in 1922 (57) in a multinomial model and later generalized by Cramér (1946), Hájek (1970), and Le Cam (1972) (58–60). In 1939, Wald (61) combined these two principles and suggested the use of minimax estimates, which involve choosing an estimator that minimizes the maximum possible loss. Following Huber's seminal work (3), M-statistics have dominated the field of parametric robust statistics for over half a century. Nonparametric methods, e.g., the Kolmogorov–Smirnov test, Mann-Whitney-Wilcoxon Test, and Hoeffding's independence test, emerged as popular alternatives to parametric methods in 1950s, as they do not make specific assumptions about the underlying distribution of the data. In 1963, Hodges and Lehmann proposed a class of robust location estimators based on the confidence bounds of rank tests (62). In REDS I, when compared to other semiparametric mean estimators with the same breakdown point, the H-L estimator was shown to be the bias-optimal choice, which aligns Devroye, and Lerasle, Lugosi, and Oliveira's conclusion that the median of means is near-optimal in terms of concentration bounds (32) as discussed. The formal study of semiparametric models was initiated by Stein (63) in 1956. Bickel, in 1982, simplified the general heuristic necessary condition proposed by Stein (63) and derived sufficient conditions for this type of problem, adaptive estimation (64). These conditions were subsequently applied to the construction of adaptive estimates (64). It has become increasingly apparent that, in robust statistics, many estimators previously called "nonparametric" are essentially semiparametric as they are partly, though not fully, characterized by some interpretable Euclidean parameters. This approach is particularly useful in situations where the data do not conform to a simple parametric distribution but still have some structure that can be exploited. In 1984, Bickel addressed the challenge of robustly estimating the parameters of a linear model while acknowledging the possibility that the model may be invalid but still within the confines of a larger model (65). He showed by carefully designing the estimators, the biases can be very small. The paradigm shift here opens up the possibility that by defining a large semiparametric model and constructing estimators simultaneously for two or more very different semiparametric/parametric models within the large semiparametric model, then even for a parametric model belongs to the large semiparametric model but not to the semiparametric/parametric models used for calibration, the performance of these estimators might still be near-optimal due to the common nature shared by the models used by the estimators. Maybe it can be named as comparametrics. Closely related topics are "mixture model" and "constraint defined model," which were generalized in Bickel, Klaassen, Ritov, and Wellner's classic semiparametric textbook (1993) (66) and the method of sieves, introduced by Grenander in 1981 (67). As the building blocks of statistics, invariant moments can reduce the overall errors of statistical results across studies and thus can enhance the replicability of the whole community (68, 69).

Methods

Methods of generating the Table 1 are summarized below, with details in the SI Text. The d values for the invariant moments of the Weibull distribution were approximated using a Monte Carlo study, with the formulae presented in Theorem **??** and **??**. The computation of I functions is summarized in Subsection **??** and further explained in the SI Text. The computation of ASABs and ASBs is described in Subsection B. The SEs and SSEs were computed by approximating the sampling distribution using 1000 pseudorandom samples for $n = 5184$ and 50 pseudorandom samples for $n = 2654208$. The impact of the bootstrap size, ranging from $n = 2.7 \times 10^2$ to $n = 2.765 \times 10^4$, on the variance of invariant moments and U-central moments was studied using the SEs and SSEs methods described above. A brute force approach was used to estimate the maximum biases of the robust estimators discussed for the five unimodal distributions. The validity of this approach is discussed in the SI Text.

Data and Software Availability. Data for Table 1 are given in SI Dataset S1-S4. All codes have been deposited in GitHub.

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).
5. M Menon, Estimation of the shape and scale parameters of the weibull distribution. *Technometrics* **5**, 175–182 (1963).
6. SD Dubey, Some percentile estimators for weibull parameters. *Technometrics* **9**, 119–129 (1967).
7. NB Marks, Estimation of weibull parameters from common percentiles. *J. applied Stat.* **32**, 17–24 (2005).
8. K Boudt, D Caliskan, C Croux, Robust explicit estimators of weibull parameters. *Metrika* **73**, 187–209 (2011).
9. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models. iii. dispersion in *Selected works of EL Lehmann.* (Springer), pp. 499–518 (2012).
10. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models iv. spread in *Selected Works of EL Lehmann.* (Springer), pp. 519–526 (2012).
11. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
12. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
13. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).
14. J SAW, Moments of sample moments of censored samples from a normal population. *Biometrika* **45**, 211–221 (1958).
15. EA Elamir, AH Seheult, Trimmed l-moments. *Comput. Stat. & Data Analysis* **43**, 299–314 (2003).
16. FR Hampel, The influence curve and its role in robust estimation. *J. american statistical association* **69**, 383–393 (1974).
17. CF Gauss, Bestimmung der genauigkeit der beobachtungen. *Ibidem* pp. 129–138 (1816).
18. H Oja, On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. statistics* pp. 154–168 (1981).
19. H Oja, Descriptive statistics for multivariate distributions. *Stat. & Probab. Lett.* **1**, 327–332 (1983).
20. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models ii. location in *selected works of EL Lehmann.* (Springer), pp. 473–497 (2012).
21. W van Zwet, Convex transformations: A new approach to skewness and kurtosis in *Selected Works of Willem van Zwet.* (Springer), pp. 3–11 (2012).
22. PJ Rousseeuw, C Croux, Alternatives to the median absolute deviation. *J. Am. Stat. association* **88**, 1273–1283 (1993).
23. PM Heffernan, Unbiased estimation of central moments by using u-statistics. *J. Royal Stat. Soc. Ser. B (Statistical Methodol.* **59**, 861–863 (1997).
24. J Hodges, E Lehmann, Matching in paired comparisons. *The Annals Math. Stat.* **25**, 787–791 (1954).
25. RA Fisher, Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.* **2**, 199–238 (1930).
26. PR Halmos, The theory of unbiased estimation. *The Annals Math. Stat.* **17**, 34–43 (1946).
27. W Hoeffding, A class of statistics with asymptotically normal distribution. *The Annals Math. Stat.* **19**, 293–325 (1948).
28. RJ Serfling, Generalized l-, m-, and r-statistics. *The Annals Stat.* **12**, 76–86 (1984).
29. E Joly, G Lugosi, Robust estimation of u-statistics. *Stoch. Process. their Appl.* **126**, 3760–3773 (2016).
30. P Laforgue, S Clémençon, P Bertail, On medians of (randomized) pairwise means in *International Conference on Machine Learning.* (PMLR), pp. 1272–1281 (2019).
31. FR Hampel, *Contributions to the theory of robust estimation.* (University of California, Berkeley), (1968).
32. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.* **44**, 2695–2725 (2016).
33. SJ Devlin, R Gnanadesikan, JR Kettenring, Robust estimation and outlier detection with correlation coefficients. *Biometrika* **62**, 531–545 (1975).
34. FR Hampel, A general qualitative definition of robustness. *The annals mathematical statistics* **42**, 1887–1896 (1971).
35. PJ Rousseeuw, FR Hampel, EM Ronchetti, WA Stahel, *Robust statistics: the approach based on influence functions.* (John Wiley & Sons), (2011).
36. DL Donoho, PJ Huber, The notion of breakdown point. *A festschrift for Erich L. Lehmann* **157184** (1983).
37. M Bieniek, Comparison of the bias of trimmed and winsorized means. *Commun. Stat. Methods* **45**, 6641–6650 (2016).
38. K Danielak, T Rychlik, Theory & methods: Exact bounds for the bias of trimmed means. *Aust. & New Zealand J. Stat.* **45**, 83–96 (2003).
39. 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).
40. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
41. T Mathieu, Concentration study of m-estimators using the influence function. *Electron. J. Stat.* **16**, 3695–3750 (2022).
42. PL Hsu, H Robbins, Complete convergence and the law of large numbers. *Proc. national academy sciences* **33**, 25–31 (1947).
43. A Kolmogorov, Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.* **4**, 83–91 (1933).
44. M Drton, H Xiao, Wald tests of singular hypotheses. *Bernoulli* **22**, 38–59 (2016).
45. NS Pillai, XL Meng, An unexpected encounter with cauchy and lévy. *The Annals Stat.* **44**, 2089–2097 (2016).
46. JE Cohen, RA Davis, G Samorodnitsky, Heavy-tailed distributions, correlations, kurtosis and taylor's law of fluctuation scaling. *Proc. Royal Soc. A* **476**, 20200610 (2020).
47. M Brown, JE Cohen, CF Tang, SCP Yam, Taylor's law of fluctuation scaling for semivariances and higher moments of heavy-tailed data. *Proc. Natl. Acad. Sci.* **118**, e2108031118 (2021).
48. WB Lindquist, ST Rachev, Taylor's law and heavy-tailed distributions. *Proc. Natl. Acad. Sci.* **118**, e2118893118 (2021).
49. G Brys, M Hubert, A Struyf, A robust measure of skewness. *J. Comput. Graph. Stat.* **13**, 996–1017 (2004).
50. JR Hosking, L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Royal Stat. Soc. Ser. B (Methodological)* **52**, 105–124 (1990).
51. AA Markov, *Wahrscheinlichkeitsrechnung.* (Teubner), (1912).
52. 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).
53. C Radhakrishna Rao, Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37**, 81–91 (1945).
54. D Blackwell, Conditional expectation and unbiased sequential estimation. *The Annals Math. Stat.* pp. 105–110 (1947).
55. EL Lehmann, H Scheffé, Completeness, similar regions, and unbiased estimation-part i in *Selected works of EL Lehmann.* (Springer), pp. 233–268 (2011).
56. EL Lehmann, H Scheffé, *Completeness, similar regions, and unbiased estimation—part II.* (Springer), (2012).
57. RA Fisher, On the mathematical foundations of theoretical statistics. *Philos. transactions Royal Soc. London. Ser. A, containing papers a mathematical or physical character* **222**, 309–368 (1922).
58. H Cramér, *Mathematical methods of statistics.* (Princeton university press) Vol. 43, (1999).
59. L LeCam, On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *The Annals Math. Stat.* **41**, 802–828 (1970).
60. J Hájek, Local asymptotic minimax and admissibility in estimation in *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability.* Vol. 1, pp. 175–194 (1972).
61. A Wald, Contributions to the theory of statistical estimation and testing hypotheses. *The Annals Math. Stat.* **10**, 299–326 (1939).
62. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
63. 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).
64. PJ Bickel, On adaptive estimation. *The Annals Stat.* **10**, 647–671 (1982).
65. P Bickel, Parametric robustness: small biases can be worthwhile. *The Annals Stat.* **12**, 864–879 (1984).
66. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparametric models.* (Springer) Vol. 4, (1993).
67. U Grenander, *Abstract Inference.* (1981).
68. JT Leek, RD Peng, Reproducible research can still be wrong: adopting a prevention approach. *Proc. Natl. Acad. Sci.* **112**, 1645–1646 (2015).
69. E National Academies of Sciences, et al., *Reproducibility and Replicability in Science.* (National Academies Press), (2019).