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 February 26, 2023

Descriptive statistics for parametric models currently rely heavily 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) but unsolved. Strictly speaking, it is unsolvable because 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 contaminated normal distribution, which has played a pre-eminent role in the later development of robust statistics. 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. All moments can be calculated from its estimated parameters. As expected, it is inadequate for the gamma, Perato, lognormal, and especially the generalized Gaussian distributions, because the logarithmic function does not produce a result for negative inputs (SI Dataset S1). Instead of minimizing the residuals, another old and interesting approach is arithmetically computing the parameters using one or more L-statistics as input values, e.g., the percentile estimators. Examples 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 clear that this class of estimators arithmetically utilizes the invariant structures of probability distributions (6, 11, 12). Maybe it can be named as I-statistics. Formally, an estimator is classified as an I-statistic if asymptotically it satisfies $\mathrm{I}\left(\mathrm{WA}_1, \cdots, \mathrm{WA}_l\right) = \left(\theta_1, \cdots, \theta_q\right)$ for the distribution it is consistent with, where WAs are weighted averages, θs are the population parameters it estimates. If the function I is solely defined through addition and/or subtraction, it is also an L-statistic. In the previous article, it is shown that quantile average is fundamental for all weighted averages. Based on the quantile function, I-statistic is naturally robust. For many parametric distributions, the quantile functions are much more elegant than the pdfs and cdfs. So I-statistics are often analytically obtainable. However, the performance of the above 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 work on semiparametric robust mean estimation, although greatly shrinking the asymptotic biases, binomial mean (BM_ϵ) is still inconsistent for any skewed distribution if $\epsilon > 0$ (if $\epsilon \to 0$, since the alternating sum of binomial coefficients is zero, $\mathrm{BM} \to \mu$). All robust location estimators commonly used are symmetric due to the universality of the symmetric distributions. One can construct an asymmetric trimmed mean that is consistent for a semiparametric class of skewed distributions. This approach was investigated previously, but it is not symmetric and therefore only suitable for some special applications (13). From semiparametric to parametric, an ideal robust location estimator would have a non-sample-dependent breakdown point (defined in Subsection **??**) and be consistent with any symmetric distribution and a skewed distribution with finite second moments. This is called an invariant mean. Based on the mean-symmetric weighted average-median inequality, the recombined mean is defined as

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

where d is for bias correction, $\mathrm{SWA}_{\epsilon,n}$ is $\mathrm{BM}_{\epsilon,n}$ in the first three Subsections, while other symmetric weighted averages can also be used in practice as long as the inequalities hold. The next theorem shows the significance of this composite estimator.

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

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,ϵ values 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 pdf of the exponential distribution is $f(x) = \lambda^{-1}e^{-\lambda^{-1}x}$, $\lambda \geq 0$, $x \geq 0$, the cdf is $F(x) = 1 - e^{-\lambda^{-1}x}$, $x \geq 0$. The quantile function 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). $rm_{d\approx 0.375,\epsilon=\frac{1}{8}}$ has 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 the estimation of central moments can be transformed into a location estimation problem by using U-statistics, the central moment kernel distributions have nice properties, and, in light of previous works, a series of sophisticated yet efficient robust estimators can be constructed whose biases are typically smaller than the variances ($n = 5400$, Table ??) 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 has the form $F(x) = F_0\left(\frac{x-\mu}{\lambda}\right)$, where F_0 is a "standard" distribution. Then, $F(x) = Q^{-1}(x) \to x = Q(p) = \lambda Q_0(p) + \mu$. So, any weighted average can be expressed as $\lambda \mathrm{W}_{estimator}(\epsilon) + \mu$, where $\mathrm{W}_{estimator}(\epsilon)$ is a function of $Q_0(p)$ according to the definition of the weighted average. The simultaneous cancellation of μ and λ in $\frac{(\lambda \mathrm{W}_\mu(\epsilon)+\mu) - (\lambda \mathrm{W}_{\mathrm{BM}_\epsilon}(\epsilon)+\mu)}{(\lambda \mathrm{W}_{\mathrm{BM}_\epsilon}(\epsilon)) + \mu - (\lambda \mathrm{W}_{median}(\epsilon)+\mu)}$ ensures 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 for the definition of invariant mean, only symmetric weighted averages are considered here.

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

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

provided that $\hat{F}_n(\mathrm{SWA}_{\epsilon,n}) \geq \frac{1}{2}$, 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, $\hat{F}_n(x) := \frac{1}{n}\left(\frac{x - \hat{Q}_n\left(\frac{sp}{n}\right)}{\hat{Q}_n\left(\frac{1}{n}(sp+1)\right) - \hat{Q}_n\left(\frac{sp}{n}\right)} + sp\right)$, where $sp = \sum_{i=1}^{n}\mathbf{1}_{X_i\leq x}$, $\mathbf{1}_A$ is the indicator of event A. The solution of $\hat{F}_n(\mathrm{SWA}_{\epsilon,n}) < \frac{1}{2}$ is reversing the percentile by $1 - \hat{F}_n(\mathrm{SWA}_{\epsilon,n})$, the obtained percentile is also reversed. Without loss of generality, in the following discussion, only the $\hat{F}_n(\mathrm{SWA}_{\epsilon,n}) \geq \frac{1}{2}$ case will be 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 happens. 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, $d = \frac{F(\mu) - F(\mathrm{BM}_\epsilon)}{F(\mathrm{BM}_\epsilon) - \frac{1}{2}}$. Recall the cdf is $F(x) = 1 - e^{-\lambda^{-1}x}$, $x \geq 0$, the expectation of BM_ϵ can be expressed as $\lambda \mathrm{W}_{\mathrm{BM}_\epsilon}(\epsilon)$, 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. $\square$

The definitions of location and scale parameters are such that they must satisfy $F(x;\lambda,\mu) = F(\frac{x-\mu}{\lambda};1,0)$. Recall that $x = \lambda Q_0(p) + \mu$, so the percentile of any weighted average is free of λ and μ, guaranteeing the validity of the quantile mean. $qm_{d\approx 0.321,\epsilon=\frac{1}{8}}$ works better in the fat-tail scenarios (SI Dataset S1). Theorem .1 and A.1 show that $rm_{d\approx 0.375,\epsilon=\frac{1}{8}}$ and $qm_{d\approx 0.321,\epsilon=\frac{1}{8}}$ 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 $rm_{d\approx 0.375,\epsilon=\frac{1}{8}}$ and $qm_{d\approx 0.321,\epsilon=\frac{1}{8}}$ 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 consider a symmetric weighted average as a mixture of symmetric quantile averages. Although using a less-biased symmetric weighted average can generally improve performance (SI Dataset S1), there is a higher risk of violation in the semiparametric framework. However, another key factor determining the risk of violation is the skewness of the distribution. Consider there is a near-symmetric distribution S such that SQA_ϵ as a function of ϵ is monotonic increasing from 0 to u and monotonic decreasing from u to $\frac{1}{2}$, and $\mu = m$. Based on the definition, S is not ordered. Depending on the fluctuation degree, $|\mathrm{SQA}_u - m|$, the mean-SWA_ϵ-median inequality is usually not valid for S. Then, the monotonic increase from 0 to u implies that $Q'(\epsilon) \geq Q'(1-\epsilon) \Leftrightarrow f(Q(1-\epsilon)) \geq f(Q(\epsilon))$ always holds for $0 \leq \epsilon \leq u$. Similarly, $Q'(\epsilon) \leq Q'(1-\epsilon) \Leftrightarrow f(Q(\epsilon)) \geq f(Q(1-\epsilon))$ always holds for $u \leq \epsilon \leq \frac{1}{2}$. Transforming S with a function $\chi(x)$ such that $\frac{d^2\chi}{dx^2} \geq 0 \wedge \frac{d\chi}{dx} \geq 0$ over the interval supported will decrease $f(Q(\epsilon))$, and the decrease rate, due to the order, is much smaller than $f(Q(1-\epsilon))$. That means, as the second derivative of $\chi(x)$ increases, eventually, after a point, $f(Q(\epsilon))$ will always be greater than $f(Q(1-\epsilon))$, i.e., the SQA_ϵ function will be monotonic decreasing and S will eventually be ordered. $|\mu - m|$ will increase so that the skewness will increase. Even the transformation is not enough to make the SQA function strictly monotonic; suppose it 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 $|u-\frac{1}{2}| \ll 1-2\epsilon$, and other parts of the SQA function satisfy the inequality constraints which define the νth orderliness, the mean-SWA_ϵ-median inequality will still be valid (as an example, 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, yet the mean-$\mathrm{BM}_{\frac{1}{8}}$-median inequality is still valid when $\alpha \leq 3.322$). Accordingly, in a family of distributions that differ by a skewness-increasing transformation, the violation of the mean-SWA_ϵ-median inequality often only occurs when the distribution is near-symmetric, but the over-corrections in rm and qm are dependent on the SWA_ϵ-median difference, which is correlated to the skewness, so the over-correction, if it happens, 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.

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).