

# Near-consistent robust estimations of moments for unimodal distributions

Tuban Lee<sup>a,1</sup>

<sup>a</sup>Institute of Biomathematics, Macau SAR 999078, China

This manuscript was compiled on March 14, 2023

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

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

The asymptotic inconsistencies between sample mean ( $\bar{x}$ ) and nonparametric robust location estimators in asymmetric distributions on the real line have been noticed for more than two centuries (1), yet remain unsolved. Strictly speaking, it is unsolvable 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 the 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, from which all moments can be calculated from its estimated parameters. As expected, it is inadequate for the gamma, Perato, lognormal, and the generalized Gaussian distributions (SI Dataset S1). Another old and interesting approach is arithmetically computing the parameters using one or more  $L$ -statistics as inputs, such as percentile estimators. Examples for the Weibull distribution, the reader is referred to Menon (1963) (6), Dubey (1967) (7), Hassanein (1971) (8), Marks (2005) (9), and Boudt, Caliskan, and Croux (2011) (10)'s works. At the outset of the study of percentile estimators, it was known that they arithmetically utilizes the invariant structures of probability distributions (6, 11, 12). Maybe such estimators can be named as  $I$ -statistics. Formally, an estimator is classified as an  $I$ -statistic if it asymptotically satisfies  $I(LE_1, \dots, LE_l) = (\theta_1, \dots, \theta_q)$  for the distribution it is consistent with, where LEs are calculated with the use of  $L$ -statistics,  $I$  is defined using arithmetic operations and constants, but it may also incorporate other functions, and  $\theta$ s are the population parameters it estimates. A subclass of  $I$ -statistics, arithmetic  $I$ -statistics, is defined as LEs are  $L$ -statistics,  $I$  is solely defined using arithmetic operations and constants. Since some percentile estimators use the logarithmic

function to transform all random variables before compute the  $L$ -statistics, a percentile estimator might not always be an arithmetic  $I$ -statistic (7). In this article, two subclasses of  $I$ -statistics are introduced, arithmetic  $I$ -statistics and quantile  $I$ -statistics. Examples of quantile  $I$ -statistics will be discussed later. Based on  $L$ -statistics,  $I$ -statistics are naturally robust. Compared to probability density functions (pdfs) and cumulative distribution functions (cdfs), the quantile functions of many parametric distributions are often more elegant. Since the expectation of an  $L$ -statistic can often be expressed as an integral of the quantile function,  $I$ -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 research on semiparametric robust mean estimation, the binomial mean ( $BM_\epsilon$ ) is still inconsistent for any skewed distribution if  $\epsilon > 0$ , although its asymptotic bias is much smaller than that of the trimmed mean (if  $\epsilon \rightarrow 0$ ,  $BM \rightarrow \mu$ , since the alternating sum of binomial coefficients is zero). All robust location estimators commonly used are symmetric due to the universality of the symmetric distributions. One can construct an asymmetric weighted average that is consistent for a semiparametric class of skewed distributions. This approach has been investigated previously, but its lack of symmetry makes it suitable only for certain applications (13). Moving from semiparametrics to parametrics, 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

## 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.

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

on the mean-symmetric weighted average-median inequality, the recombined mean is defined as

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

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

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

*Proof.* Finding  $d$  and  $\epsilon$  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,  $rm_{d,\epsilon} = \lim_{c \rightarrow \infty} \left( \frac{(\text{BM}_{\epsilon} + c)^{d+1}}{(\text{median} + c)^d} - c \right) = (d+1)\text{BM}_{\epsilon} - d\text{median} = \mu$ . So,  $d = \frac{\mu - \text{BM}_{\epsilon}}{\text{BM}_{\epsilon} - \text{median}}$ . The quantile function of the exponential distribution is  $Q(p) = \ln\left(\frac{1}{1-p}\right)\lambda$ .  $E[x] = \lambda$ .  $E[\text{median}] = Q\left(\frac{1}{2}\right) = \ln 2\lambda$ . For the exponential distribution, the expectation of  $\text{BM}_{\frac{1}{8}}$  is  $E\left[\text{BM}_{\frac{1}{8}}\right] = \lambda \left(1 + \ln\left(\frac{46656}{8575\sqrt{35}}\right)\right)$ . Obviously, the scale parameter  $\lambda$  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,  $E[\text{BM}_{\epsilon}] = E[\text{median}] = E[X]$ . Then  $E[rm_{d,\epsilon}] = \lim_{c \rightarrow \infty} \left( \frac{(E[X] + c)^{d+1}}{(E[X] + c)^d} - c \right) = E[X]$ . The proof for the Pareto distribution is more general. The mean of the Pareto distribution is given by  $\frac{\alpha x_m}{\alpha - 1}$ . The  $d$  value with two unknown percentiles  $p_1$  and  $p_2$  for the Pareto distribution is  $d_{\text{Pareto}} = \frac{\frac{\alpha x_m}{\alpha - 1} - x_m(1-p_1)^{-\frac{1}{\alpha}}}{x_m(1-p_1)^{-\frac{1}{\alpha}} - x_m(1-p_2)^{-\frac{1}{\alpha}}}$ . Since any weighted average can be expressed as an integral of the quantile function,  $\lim_{\alpha \rightarrow \infty} \frac{\frac{\alpha}{\alpha-1} - (1-p_1)^{-1/\alpha}}{(1-p_1)^{-1/\alpha} - (1-p_2)^{-1/\alpha}} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$ , the  $d$  value for the Pareto distribution approaches that of the exponential distribution as  $\alpha \rightarrow \infty$ , regardless of the type of weighted 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, in light of previous works, the estimation of central moments can be transformed into a location estimation problem by using  $U$ -statistics, the central moment kernel distributions possess desirable properties, and a series of sophisticated yet efficient robust estimators can be constructed whose biases are typically smaller than the variances (as seen in Table ?? for  $n = 5400$ ) for unimodal distributions.

## Background and Main Results

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

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

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

provided that  $\hat{F}_n(\text{SWA}_{\epsilon,n}) \geq \frac{1}{2}$ , 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,  $\hat{F}_n(x) := \frac{1}{n} \left( \frac{x - X_{sp}}{X_{sp+1} - X_{sp}} + sp \right)$ , where  $sp = \sum_{i=1}^n \mathbf{1}_{X_i \leq x}$ ,  $\mathbf{1}_A$  is the indicator of event  $A$ . The solution of  $\hat{F}_n(\text{SWA}_{\epsilon,n}) < \frac{1}{2}$  is reversing the percentile by  $1 - \hat{F}_n(\text{SWA}_{\epsilon,n})$ , the obtained percentile is also reversed. Without loss of generality, in the following discussion, only the case where  $\hat{F}_n(\text{SWA}_{\epsilon,n}) \geq \frac{1}{2}$  is considered. Moreover, in extreme heavy-tailed distributions, the calculated percentile can exceed the breakdown point of  $\text{SWA}_{\epsilon}$ , so the percentile will be modified to  $1 - \epsilon$  if this occurs. The quantile mean uses the location-scale invariant in a different way as shown in the following proof.

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

*Proof.* Similarly, rearranging the definition,  $d = \frac{F(\mu) - F(\text{BM}_{\epsilon})}{F(\text{BM}_{\epsilon}) - \frac{1}{2}}$ . The cdf of the exponential distribution is  $F(x) = 1 - e^{-\lambda^{-1}x}$ ,  $\lambda \geq 0$ ,  $x \geq 0$ , the expectation of  $\text{BM}_{\epsilon}$  can be expressed as  $\lambda \text{BM}_0(\epsilon)$ , so  $F(\text{BM}_{\epsilon})$  is free of  $\lambda$ . 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,  $F(E[\text{BM}_{\epsilon}]) = F(\mu) = \frac{1}{2}$ . Then, the expectation of the quantile mean is  $qm_{d,\epsilon} = F^{-1}\left(\left(F(\mu) - \frac{1}{2}\right)d + F(\mu)\right) = F^{-1}\left(0 + F(\mu)\right) = \mu$ .

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

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

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

are now proven.  $\square$

The definitions of location and scale parameters are such that they must satisfy  $F(x; \lambda, \mu) = F\left(\frac{x-\mu}{\lambda}; 1, 0\right)$ . Recall that  $x = \lambda Q_0(p) + \mu$ , so the percentile of any weighted average is free of  $\lambda$  and  $\mu$ , which guarantees the validity of the quantile mean. The quantile mean is a quantile  $I$ -statistic. Specifically, an estimator is classified as a quantile  $I$ -statistic if LEs are percentiles of a distribution obtained by plugging  $L$ -statistics into a cumulative distribution function and  $I$  is defined with arithmetic operations, constants and quantile functions.  $qm_{d \approx 0.321, \epsilon = \frac{1}{8}}$  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 evident 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 recall that a symmetric weighted average is a linear combination of symmetric quantile averages. While using a less-biased symmetric weighted average can generally enhance performance (SI Dataset S1), there is a greater risk of violation in the semiparametric framework. However, the mean-SWA-median inequality is robust to slight fluctuations of the SQA function of the underlying distribution. Suppose the SQA function is generally decreasing in  $[0, u]$ , but increasing in  $[u, \frac{1}{2}]$ , since  $1-2\epsilon$  of the symmetric quantile averages will be included in the computation of  $SWA_\epsilon$ , as long as  $|u - \frac{1}{2}| \ll 1-2\epsilon$ , and other portions of the SQA function satisfy the inequality constraints that define the  $\nu$ th orderliness on which the  $SWA_\epsilon$  is based, the mean- $SWA_\epsilon$ -median inequality will still hold. This is due to the violation being bounded (15) and therefore cannot be extreme for unimodal distributions. For instance, the SQA function is non-monotonic when the shape parameter of the Weibull distribution  $\alpha > \frac{1}{1-\ln(2)} \approx 3.259$  as shown in the previous article, the violation of the third orderliness is also close to this parameter, yet the mean- $BM_{\frac{1}{8}}$ -median inequality is still valid when  $\alpha \leq 3.322$ . Another key factor in determining the risk of violation is the skewness of the distribution. In the previous article, it was demonstrated that in a family of distributions differing by a skewness-increasing transformation in van Zwet's sense, the violation of orderliness, if it happens, often only occurs when the distribution is nearly symmetrical (16). The over-corrections in  $rm$  and  $qm$  are dependent on the  $SWA_\epsilon$ -median difference, which can be a reasonable measure of skewness (17, 18), implying that the over-correction is often tiny with a moderate  $d$ . This qualitative analysis provides another perspective, in addition to the bias bounds (15), that  $rm$  and  $qm$  based on the mean- $SWA_\epsilon$ -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). 210
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). 212
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**, 1 (1912). 213
4. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964). 214
5. X He, WK Fung, Method of medians for lifetime data with weibull models. *Stat. medicine* **18**, 1993–2009 (1999). 215
6. M Menon, Estimation of the shape and scale parameters of the weibull distribution. *Technometrics* **5**, 175–182 (1963). 216
7. SD Dubey, Some percentile estimators for weibull parameters. *Technometrics* **9**, 119–129 (1967). 217
8. KM Hassanein, Percentile estimators for the parameters of the weibull distribution. *Biometrika* **58**, 673–676 (1971). 218
9. NB Marks, Estimation of weibull parameters from common percentiles. *J. applied Stat.* **32**, 17–24 (2005). 219
10. K Boudt, D Caliskan, C Croux, Robust explicit estimators of weibull parameters. *Metrika* **73**, 187–209 (2011). 220
11. SD Dubey, *Contributions to statistical theory of life testing and reliability*. (Michigan State University of Agriculture and Applied Science. Department of statistics), (1960). 221
12. LJ Bain, CE Antle, Estimation of parameters in the weibull distribution. *Technometrics* **9**, 621–627 (1967). 222
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). 223
14. RJ Hyndman, Y Fan, Sample quantiles in statistical packages. *The Am. Stat.* **50**, 361–365 (1996). 224
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). 225
16. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen*. (1964). 226
17. AL Bowley, *Elements of statistics*. (King) No. 8, (1926). 227
18. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.)* **33**, 391–399 (1984). 228