Near-consistent robust estimations of moments for unimodal distributions

Tuban Lee

This manuscript was compiled on May 26, 2023

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

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

The potential inconsistencies between the sample mean ($\bar{x}$) and robust location estimators in distributions with finite moments have been noticed for more than two centuries (1), with numerous significant attempts made to address them. In calculating a robust location estimator, the procedure of identifying and downweighting extreme values inherently necessitates the formulation of certain distributional assumptions. Inconsistencies natually arise when these assumptions, parametric or semiparametric, are violated. Due to the presence of infinite dimensional nuisance shape parameters, the semiparametric approach struggles to adequately address distributions with more intricate shapes. Newcomb (1886) provided the first modern approach to robust parametric estimation by developing a class of estimators that gives "less weight to the more discordant observations" (2). In 1964, Huber (3) used the minimax procedure to obtain M-estimator for the contaminated normal distribution, which has played a pre-eminent role in the later development of robust statistics. However, as previously demonstrated, under growing asymmetric departures from normality, the bias of the Huber M-estimator increases rapidly. This is a common issue in parametric robust statistics. For example, He and Fung (1999) constructed (4) a robust M-estimator for the two-parameter Weibull distribution, from which all moments can be calculated. Nonetheless, it is inadequate for 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 probability distributions (5, 6). Maybe such estimators can be named as I-statistics. Formally, an estimator is classified as an I-statistic if it asymptotically satisfies $\mathrm{I}\left(\mathrm{LE}_1, \ldots, \mathrm{LE}_l\right) = (\theta_1, \ldots, \theta_q)$ for the distribution it is consistent, where LEs are calculated with the use of LU-statistics (defined in Subsection **??**), 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. A subclass of I-statistics, arithmetic I-statistics, is defined as LEs are LU-statistics, I is solely defined using arithmetic operations and constants. Since some percentile estimators use the logarithmic function to transform all random variables before computing the L-estimators, a percentile estimator might not always be an arithmetic I-statistic (6). In this article, two subclasses of I-statistics are introduced, arithmetic I-statistics and quantile I-statistics. Examples of quantile I-statistics will be discussed later. Based on LU-statistics, I-statistics are naturally robust. Compared to probability density functions (pdfs) and cumulative distribution functions (cdfs), the quantile functions of many parametric distributions are more elegant. Since the expectation of an L-estimator can be expressed as an integral of the quantile function, I-statistics are often analytically obtainable. However, it is observed that 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 (7), rendering the approach ill-suited (SI Dataset S1).

DRAFT

Most robust location estimators commonly used are symmetric owing to the prevalence of symmetric distributions. An asymmetric γ-weighted L-statistic can achieve consistency for any γ-symmetric distribution, if $\gamma \neq 1$. However, it is tailored more towards certain specific distributions rather than a broad spectrum of common ones. Shifting from semiparametrics to parametrics, consider an estimator with a non-sample-dependent breakdown point (defined in Subsection **??**) that is consistent simultanously for both a semiparametric class of distributions and a distinct parametric distribution, such a robust estimator is named with the prefix 'invariant' followed by the population parameter it is consistent with. Here, the recombined mean is defined as

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

where d is the key factor for bias correction, γm_n is the sample γ-median, $\mathrm{WL}_{\epsilon,\gamma,n}$ is the weighted L-statistic. It is assumed

Significance Statement

Bias, variance, and contamination are the three main errors in statistics. Consistent robust estimation is unattainable without parametric assumptions. In this article, invariant moments are proposed as a means of achieving near-consistent and robust estimations of moments, even in scenarios where moderate violations of distributional assumptions occur, while the variances are sometimes smaller than those of the sample moments.

T.L. designed research, performed research, analyzed data, and wrote the paper.

The author declares no competing interest.

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

in this article that in the subscript of an invariant moment, if γ is omitted, $\gamma = 1$ is assumed, if n is omitted, only the asymptotic behavior is considered. The subsequent theorem shows the significance of this arithmetic I-statistic.

Theorem .1. *$rm_{d\approx 0.103,\nu=3,\epsilon=\frac{1}{24},BM}$ is a consistent mean estimator for the exponential distribution, any symmetric distributions and the Pareto distribution with quantile function $Q(p) = x_m(1-p)^{-\frac{1}{\alpha}}$, $x_m > 0$, when $\alpha \to \infty$, provided that the second moments are finite.*

Proof. Finding d that make $rm_{d,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ a consistent mean estimator is equivalent to finding the solution of $rm_{d,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}} = \mu$. First consider the exponential distribution, whose quantile function is $Q(p) = \ln\left(\frac{1}{1-p}\right)\lambda$. $\mu = \lambda$. $m = Q\left(\frac{1}{2}\right) = \ln 2\lambda$. $\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}} = \lambda\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)$, the detailed formula is given in the SI Text. Since $rm_{d,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}} = \lim_{c\to\infty}\left(\frac{\left(\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}+c\right)^{d+1}}{(m+c)^d} - c\right) = (d+1)\,\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}} - dm = \mu$. So, $d = \frac{\mu-\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}}{\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}-m} = \frac{\lambda-\lambda\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)}{\lambda\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)-\ln 2\lambda} = -\frac{\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)}{1-\ln(2)+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)} \approx 0.103$. The proof of the second assertion follows directly from the coincidence property. According to Theorem 20 in the previous article, for any γ-symmetric distribution with a finite second moment, $\mathrm{WL}_{\epsilon,\gamma} = Q(\frac{\gamma}{1+\gamma}) = \mu$. Then $rm_{d,\epsilon,\gamma,\mathrm{WL}} = \lim_{c\to\infty}\left(\frac{(\mu+c)^{d+1}}{(\mu+c)^d} - c\right) = \mu$. The proof for the Pareto distribution is more general. The mean of the Pareto distribution is given by $\frac{\alpha x_m}{\alpha-1}$. Since any weighted L-statistic can be expressed as an integral of the quantile function as shown previously, the γ-median is also a percentile, one can replace the WL and γm in the d value with two arbitrary percentiles p_1 and p_2. For the Pareto distribution, $d_{Perato} = \frac{\mu-Q(p_1)}{Q(p_1)-Q(p_2)} = \frac{\frac{\alpha x_m}{\alpha-1}-x_m(1-p_1)^{-\frac{1}{\alpha}}}{x_m(1-p_1)^{-\frac{1}{\alpha}}-x_m(1-p_2)^{-\frac{1}{\alpha}}}$. x_m can be canceled out. For the exponential distribution, $d_{exp} = \frac{\mu-Q(p_1)}{Q(p_1)-Q(p_2)} = \frac{\lambda-\ln\left(\frac{1}{1-p_1}\right)\lambda}{\ln\left(\frac{1}{1-p_1}\right)\lambda-\ln\left(\frac{1}{1-p_2}\right)\lambda} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$. Since $\lim_{\alpha\to\infty}\frac{\frac{\alpha}{\alpha-1}-(1-p_1)^{-1/\alpha}}{(1-p_1)^{-1/\alpha}-(1-p_2)^{-1/\alpha}} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$, the d value for the Pareto distribution approaches that of the exponential distribution, as $\alpha \to \infty$, regardless of the type of weighted L-statistic used. This completes the demonstration. □

Theorem .1 implies that for the Weibull, gamma, Pareto, lognormal and generalized Gaussian distribution, $rm_{d\approx 0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ is consistent for at least one particular case. The biases of $rm_{d\approx 0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ for distributions with skewness between those of the exponential and symmetric distributions are tiny (SI Dataset S1). $rm_{d\approx 0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ exhibits excellent performance for all these common unimodal distributions (SI Dataset S1).

Besides introducing the concept of invariant mean, the purpose of this paper is to demonstrate that, in light of previous works, the estimation of central moments can be transformed into a location estimation problem by using U-statistics, the central moment kernel distributions possess desirable properties, and a series of sophisticated yet efficient robust estimators can be constructed whose biases are typically smaller than the variances (as seen in Table **??** for $n = 4096$) for unimodal distributions.

A. Invariant mean. Previously, it was established that any $\mathrm{WL}(\epsilon,\gamma)$ can be expressed as $\lambda\mathrm{WL}_0(\epsilon,\gamma)+\mu$ for a location-scale distribution parameterized by a location parameter μ and a scale parameter λ, where WL_0 denote the expected value of the weighted L-statistic for a standard distribution without any shifts or scaling. The simultaneous cancellation of μ and λ in $\frac{(\lambda\mu_0+\mu)-(\lambda\mathrm{WL}_0(\epsilon,\gamma)+\mu)}{(\lambda\mathrm{WL}_0(\epsilon,\gamma)+\mu)-(\lambda\gamma m_0+\mu)}$ assures that the d in rm is always a constant for a location-scale distribution.

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

$$qm_{d,\epsilon,\gamma,n,\mathrm{WL}} \coloneqq \hat{Q}_n\left(\left(\hat{F}_n\left(\mathrm{WL}_{\epsilon,\gamma,n}\right)-\frac{1}{1+\gamma}\right)d+\hat{F}_n\left(\mathrm{WL}_{\epsilon,\gamma,n}\right)\right),$$

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

Theorem A.1. *$qm_{d\approx 0.088,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ is a consistent mean estimator for the exponential, Pareto ($\alpha\to\infty$) and any symmetric distributions provided that the second moments are finite.*

Proof. The cdf of the exponential distribution is $F(x) = 1-e^{-\lambda^{-1}x}$, $\lambda \geq 0$, $x \geq 0$. Recall that $\mathrm{BM}_{\nu=3,\epsilon}$ can be expressed as $\lambda\mathrm{BM}_0(\epsilon)$, so $F(\mathrm{BM}_{\nu=3,\epsilon})$ is free of λ, as is $F(\mu)$. When $\epsilon = \frac{1}{24}$, $d = \frac{F(\mu)-F(\mathrm{BM}_{\nu=3,\epsilon})}{F(\mathrm{BM}_{\nu=3,\epsilon})-\frac{1}{2}} = \frac{-e^{-1}+e^{-\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)}}{\frac{1}{2}-e^{-\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)}} = \frac{\frac{101898752449325\sqrt{5}\sqrt[6]{\frac{247}{7}}391^{5/6}}{26068394603446272\sqrt[3]{11}e}-\frac{1}{e}}{\frac{1}{2}-\frac{101898752449325\sqrt{5}\sqrt[6]{\frac{247}{7}}391^{5/6}}{26068394603446272\sqrt[3]{11}e}} \approx 0.088$. The proof of the symmetric case: since for any γ-symmetric distribution

DRAFT

with a finite second moment, $F\left(\mathrm{WL}_{\epsilon,\gamma}\right)=F\left(\mu\right)=\frac{\gamma}{1+\gamma}$. Then, the expectation of the quantile mean is $qm_{d,\epsilon,\gamma,\mathrm{WL}}=F^{-1}\left(\left(F\left(\mathrm{WL}_{\epsilon,\gamma}\right)-\frac{\gamma}{1+\gamma}\right)d+F\left(\mu\right)\right)=F^{-1}\left(0+F\left(\mu\right)\right)=\mu$.

For the assertion related to the Pareto distribution, the cdf of it is $1-\left(\frac{x_m}{x}\right)^{\alpha}$. Similar to Theorem .1, replacing the $F(\mathrm{WL}_{\epsilon,\gamma})$ and $\frac{1}{1+\gamma}$ in the d value with two arbitrary percentiles p_1 and p_2,

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

When $\alpha\rightarrow\infty$, $\left(\frac{\alpha-1}{\alpha}\right)^{\alpha}=\frac{1}{e}$, so in this case, d_{Pareto} is identical to that of the exponential distribution, since

$$d_{exp}=\frac{\left(1-e^{-1}\right)-\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)}{\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)-\left(1-e^{-\ln\left(\frac{1}{1-p_2}\right)}\right)}=\frac{1-\frac{1}{e}-p_1}{p_1-p_2}.$$

Therefore, same logic as in Theorem .1, their d values are always identical, regardless of the type of weighted L-statistic used. All results are now proven. □

The definitions of location and scale parameters are such that they must satisfy $F(x;\lambda,\mu)=F(\frac{x-\mu}{\lambda};1,0)$. By recalling $\mathrm{WL}=\lambda\mathrm{WL}_0(\epsilon,\gamma)+\mu$, it follows that the percentile of any weighted L-statistic is free of λ and μ, which guarantees the validity of the quantile mean. The quantile mean is a quantile I-statistic. Specifically, an estimator is classified as a quantile I-statistic if LEs are percentiles of a distribution obtained by plugging LU-statistics into a cumulative distribution function and I is defined with arithmetic operations, constants and quantile functions. $qm_{d\approx0.088,\nu=3,\epsilon=\frac{1}{24},\mathrm{WL}}$ works better in the fat-tail scenarios (SI Dataset S1). Theorem .1 and A.1 show that $rm_{d\approx0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ and $qm_{d\approx0.088,\nu=3,\epsilon=\frac{1}{24},\mathrm{WL}}$ are both consistent mean estimators for any symmetric distribution and a skewed distribution with finite second moments. The breakdown points of $rm_{d\approx0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ and $qm_{d\approx0.088,\nu=3,\epsilon=\frac{1}{24},\mathrm{WL}}$ are both $\frac{1}{24}$. Therefore they are all invariant means.

To study the impact of the choice of WLs in rm and qm, it is constructive to recall that a weighted L-statistic is a combination of order statistics. While using a less-biased weighted L-statistic can generally enhance performance (SI Dataset S1), there is a greater risk of violation in the semiparametric framework. However, the mean-$\mathrm{WA}_{\epsilon,\gamma}$-$\gamma$-median inequality is robust to slight fluctuations of the QA function of the underlying distribution when $0\leq\gamma\leq1$. Suppose the QA function is generally decreasing in $[0,u]$, but increasing in $[u,\frac{1}{1+\gamma}]$, since all quantile averages with breakdown points from ϵ to $\frac{1}{1+\gamma}$ will be included in the computation of $\mathrm{WA}_{\epsilon,\gamma}$, as long as $\frac{1}{1+\gamma}-u\ll\frac{1}{1+\gamma}-\gamma\epsilon$, and other portions of the QA function satisfy the inequality constraints that define the νth γ-orderliness on which the $\mathrm{WA}_{\epsilon,\gamma}$ is based, the mean-$\mathrm{WA}_{\epsilon,\gamma}$-$\gamma$-median inequality still holds. This is due to the violation being bounded (9) when $0\leq\gamma\leq1$ and therefore cannot be extreme for unimodal distributions with finite second moments. For instance, the SQA function is non-monotonic when the shape parameter of the Weibull distribution $\alpha>\frac{1}{1-\ln(2)}\approx3.259$ as shown in the previous article, the violation of the third orderliness starts near this parameter as well, yet the mean-$\mathrm{BM}_{\frac{1}{24}}$-median inequality retains valid when $\alpha\leq3.387$. Another key factor in determining the risk of violation is the skewness of the distribution. Previously, it was demonstrated that in a family of distributions differing by a skewness-increasing transformation in van Zwet's sense, the violation of orderliness, if it happens, often only occurs as the distribution nears symmetry (10). When $\gamma=1$, the over-corrections in rm and qm are dependent on the SWA_{ϵ}-median difference, which can be a reasonable measure of skewness after standardization (11, 12), implying that the over-correction is often tiny with moderate d. This qualitative analysis suggests the general reliability of rm and qm based on the mean-$\mathrm{WA}_{\epsilon,\gamma}$-$\gamma$-median inequality for unimodal distributions with finite second moments. Extending this rationale to other weighted L-statistics is possible, since the γ-U-orderliness can also be bounded with certain assumptions, as discussed previously.

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

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

1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae*. (Henricus Dieterich), (1823).
2. S Newcomb, A generalized theory of the combination of observations so as to obtain the best result. *Am. journal Math.* **8**, 343–366 (1886).
3. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
4. X He, WK Fung, Method of medians for lifetime data with weibull models. *Stat. medicine* **18**, 1993–2009 (1999).
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. C Bernard, R Kazzi, S Vanduffel, Range value-at-risk bounds for unimodal distributions under partial information. *Insur. Math. Econ.* **94**, 9–24 (2020).
10. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen*. (1964).
11. AL Bowley, *Elements of statistics*. (King) No. 8, (1926).
12. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).

DRAFT