Near-consistent robust estimations of moments for unimodal distributions

Tuban Lee

This manuscript was compiled on May 31, 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. 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 γ-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 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). 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 (7), rendering the approach ill-suited (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). Throughout this paper, it is assumed that all underlying distributions have finite moments and are continuous. 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 ?? for $n = 4096$).

A. Robust Estimations of the Central Moments. In 1979, Bickel and Lehmann (10), in their final paper of the landmark series *Descriptive Statistics for Nonparametric Models*, generalized a class of estimators called measures of spread, which "do not require the assumption of symmetry." From this, a popular efficient scale estimator, the Rousseeuw-Croux scale estimator (16), was derived in 1993, but the importance of tackling the symmetry assumption has been greatly underestimated. While they had already considered one version of the trimmed standard deviation, which is a measures of dispersion, in the third

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

paper of that series (9); in the final section of that paper (10), they explored another two versions of the trimmed standard deviation based on pairwise differences, one is modified here for comparison,

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

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

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

Nomenclature. Given a robust estimator $\hat{\theta}$, which has an adjustable breakdown point that can approach zero asymptotically, the name of $\hat{\theta}$ comprises two parts: the first part denotes the type of estimator, and the second part is the name of the population parameter θ, with which the estimator approaches as $\epsilon \to 0$. The abbreviation of the estimator combines the initial letters of the first part and the population parameter. If the estimator is symmetric, the upper asymptotic breakdown point, ϵ (defined in Subsection **??**, or $\epsilon_{U_\mathbf{k}}$ for LU-statistic), is indicated in the subscript of the abbreviation of the estimator, with the exception of the median. For an asymmetric estimator based on quantile average, the associated γ follows ϵ.

Proof. □

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. 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. PJ Rousseeuw, C Croux, Alternatives to the median absolute deviation. *J. Am. Stat. association* **88**, 1273–1283 (1993).