MAD (about median) vs. quantile-based alternatives for classical standard deviation, skewness, and kurtosis

Eugene Pinsky[1]* and Sidney Klawansky[2]

[1]Department of Computer Science, Metropolitan College, Boston University, Boston, MA, United States,
[2]Department of Health Policy and Management, Harvard School of Public Health, Boston, MA,
United States

In classical probability and statistics, one computes many measures of interest from mean and standard deviation. However, mean, and especially standard deviation, are overly sensitive to outliers. One way to address this sensitivity is by considering alternative metrics for deviation, skewness, and kurtosis using mean absolute deviations from the median (MAD). We show that the proposed measures can be computed in terms of the sub-means of the appropriate left and right sub-ranges. They can be interpreted in terms of average distances of values of these sub-ranges from their respective medians. We emphasize that these measures utilize only the first-order moment within each sub-range and, in addition, are invariant to translation or scaling. The obtained formulas are similar to the quantile measures of deviation, skewness, and kurtosis but involve computing sub-means as opposed to quantiles. While the classical skewness can be unbounded, both the MAD-based and quantile skewness always lies in the range $[-1, 1]$. In addition, while both the classical kurtosis and quantile-based kurtosis can be unbounded, the proposed MAD-based alternative for kurtosis lies in the range $[0, 1]$. We present a detailed comparison of MAD-based, quantile-based, and classical metrics for the six well-known theoretical distributions considered. We illustrate the practical utility of MAD-based metrics by considering the theoretical properties of the Pareto distribution with high concentrations of density in the upper tail, as might apply to the analysis of wealth and income. In summary, the proposed MAD-based alternatives provide a universal scale to compare deviation, skewness, and kurtosis across different distributions.

1. Introduction

Classical statistics uses the standard deviation σ as the primary measure of dispersion. In computing σ, we use the squares of the distances from the mean μ. As noted in [1], using the L_2 norm is convenient in differentiation, estimation, and optimization. The additive property of variance σ^2 for independent variables is also cited as one of the prime reasons for using the L_2 norm in sampling theory and analysis of variance. A historical survey is given in [2].

At the same time, this norm has a number of disadvantages. For example, large deviations from outliers contribute heavily to mean and standard deviation and could significantly overestimate "typical" deviations. A natural alternative is to use the L_1 norm and measure absolute deviations from a central point such as the mean or median.

The idea of using the L_1 norm is not new. The L_1 norm was considered independently by both Boscovitch and Laplace as early as the eighteenth century. A historical survey using the L_1 norm is presented in [3, 4] and a survey of more recent results is given in [1]. However, the L_1 norm has not been widely used in statistics and statistical modeling [1].

There is currently a renewed interest in using the L_1 norm for robust statistical modeling and inference [e.g., [5–9]]. Using the L_2 norm, the influence of outliers is even more magnified when computing skewness and kurtosis as these computations would involve raising means and standard deviation to 3rd and 4th powers. By contrast, using mean absolute deviations from the mean or median (both denoted as MAD in literature) can be more appropriate. Therefore, using the L_1 norm, outliers will have less influence on the results. Consequently, results from using the MAD (mean absolute deviation) are more robust to outliers than those obtained using the standard deviation, as is common in classical statistics.

Throughout this paper, we will use MAD to denote the mean absolute deviation from the median. We will use the MAD in deriving alternative expressions for skewness and kurtosis. The proposed MAD-based measures can be computed and interpreted as sub-means of the appropriate left and right sub-ranges. The obtained formulas are analogous to those used in statistics based on quantiles.

The MAD-based alternative measures considered in this paper use only the first-order moment. Consequently, they do not overweight outliers as compared to classical measures. When compared to the quantile metrics, they are more sensitive to concentrations at the extreme ends of distributions. In addition, these alternative measures have many desirable characteristics, such as scale (by absolute value) and/or shift-invariance.

We illustrate our results with several examples. One of the novel contributions of this paper is the MAD-based alternative metric for kurtosis. This metric is shown to be in the [0, 1] range, allowing us to compare the data distribution with different numerical scales. We present a detailed analysis of some well-known distributions. We contrast the proposed MAD-based metrics with both quantile and classical statistics metrics.

2. Organization of the paper

This paper is organized as follows. Section 3 introduces notation and reviews some bounds on mean absolute deviation. In Section 4, we focus on computing MAD, show how it can be computed as a difference of corresponding sub-means, and contrast the expression for MAD deviation with that of quantile deviation. In Section 5, we introduce MAD-based skewness and kurtosis. The MAD expression for skewness has been known before. To our knowledge, the proposed expression for MAD-based kurtosis has not been considered before. Both MAD-based skewness and kurtosis have simple interpretations, and their formulas are similar to quantile-based alternatives skewness and kurtosis. Section 6 discusses the advantages of proposed measures vs. corresponding classical and quantile-based measures. Section 7 focuses on computational considerations and shows how MAD-based measures can be computed from some integrals related to

the underlying probability distributions. In Section 8, we provide a detailed comparison for a number of distributions:

1. Continuous uniform (Section 8.1)
2. Normal (Section 8.2)
3. Log-normal (Section 8.3)
4. Exponential (Section 8.4)
5. Laplace (Section 8.5)
6. Pareto (section 8.6).

In Section 9, we present an example of applying our results to analyze wealth distribution. In Section 10, we directly compare MAD-based skewness and kurtosis for the above six distributions. We conclude our paper with Section 11.

For completeness and clarity of presentation, we moved some details of derivations into Appendices. In Appendix 1 (Section A1), we present summary tables for the above distributions. In Appendices 2–4 (Sections A2–A4) we present some computational details for log-normal, Laplace and Pareto distributions, respectively.

3. Preliminaries

We start with preliminary definitions. Consider a real-valued random variable X on a sample space $\Omega \subseteq R$ with density $f(x)$, finite mean $E(X)$, and cumulative distribution function $F(x)$. If X is a discrete random variable, then Ω is some countable sample space, and $f(x)$ is the probability mass function (discrete density function).

We use $\mu = E(X)$ and σ to denote mean and standard deviation of X. Let F^{-1} be the quantile function defined by $F^{-1}(t) = \inf\{x : F(x) \geq t\}$ with $t \in [0, 1]$. The median M, the quartiles Q_1 and Q_3 are given by $M = F^{-1}(1/2)$, $Q_1 = F^{-1}(1/4)$, and $Q_3 = F^{-1}(3/4)$, respectively.

For any a, we define the mean absolute deviation of X from a as

$$H(X, a) = E(|X - a|) = \int_\Omega |x - a| f(x)\, dx \qquad (1)$$

If $a = \mu$, then $H(X, \mu)$ is the mean absolute deviation from the mean μ. If we take $a = M$, then $H(X, M)$ is the mean absolute deviation from the median. Both of these are denoted as MAD (mean absolute deviation) in the statistical literature, leading to some confusion [1]. In this paper, we use MAD to denote **mean absolute deviations from the median**. It can be interpreted as the average distance of values of X to the median M. We will write H as an abbreviation to $H(X, M)$.

Let us start by establishing a lower and upper bound for H. Since $f(x) \geq 0$, is integrable and $E(X) < \infty$ we have $-|x - M| f(x) \leq (x - M) f(x) \leq |x - M| f(x)$ and, therefore, we obtain

$$H = \int_\Omega |x - M| f(x)\, dx \geq \left| \int_\Omega (x - M) f(x)\, dx \right| = |M - \mu|$$

To establish an upper bound, we use the well-known fact that $H \leq H(X, a)$ for any value of a [10–12]. In particular, if $a = \mu$ then $H \leq E(|X - \mu|)$. This means that the average absolute deviation from the median H is always less than or equal to the mean absolute deviation from the mean $E(|X - \mu|)$. If we apply Jensen's inequality

$E(g(X)) \geq g(\mu)$ to the convex function $g(t) = t^2$ (corresponding to σ^2) we immediately obtain an upper bound for H:

$$\sigma^2 = E\left((X - \mu)^2\right) = E\left(g(|X - \mu|)\right) \geq g\left[E(|X - \mu|)\right]$$
$$= \left[E(|X - \mu|)\right]^2 \geq H^2$$

Of the three metrics to measure deviations, namely H, $E(|X - \mu|)$ and σ, the MAD metric H has the lowest value.

Example 1: Consider two uniform random variables X and Y with corresponding sample spaces Ω_x and Ω_y with $n = 12$ elements given by

$$\Omega_X = \left\{ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 \right\}, \quad \text{and}$$
$$\Omega_Y = \left\{ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 100 \right\}$$

and the discrete density functions $f_x(\cdot) = f_y(\cdot) = 1/n$ for any value in Ω_X and Ω_Y, respectively. For variable X we have $M_x = 6.5$, $\mu_x = 6.5$, $\sigma_x = 3.45$, and $H_x = 3$ whereas for variable Y we have $M_y = 6.5$, $\mu_y = 13.83$, $\sigma_y = 26.16$, and $H_y = 10.33$. For the purpose of illustration, let us define an outlier as any numeric value v with $|v - \mu| > 2\sigma$. Both random variables X and Y have the same median $M_x = M_y = 6.5$, but the random variable Y has a much higher outlier value $y_{12} = 100$. This much higher outlier value results in a higher mean and a higher standard deviation for Y than for X. The impact of this outlier value on H can be immediately computed: it increases H from $H_x = 3$ to $H_y = 10.33$ by $(y_{12} - x_{12})/n = 7.33$. The change in standard deviations due to this outlier would involve a much more complicated expression. The MAD-based deviation H_Y for Y is about four times higher than the MAD-based deviation H_x for X, whereas the standard deviation σ_y for Y, is more than seven times greater than σ_x for X because of the squaring of the deviations. To assess the impact of the outlier $y_{12} = 100$, we compare H_x and H_y in terms of σ_x and σ_y. For variable X we have $H_x \approx 0.87\sigma_x$ whereas for variable Y we have $H_y \approx 0.39\sigma_y$. The lower value of H_y as compared with σ_y for Y indicates that Y has heavier tails compared to X.

The above example with an outlier illustrates one of the advantages of using MAD instead of standard deviation as a measure of variability. When computing the standard deviation or variance of X, the outlier effect is amplified since we square the differences $(x_i - \mu)$. This effect of outliers is further amplified in computing skewness or kurtosis where we need to raise these differences to the 3$^{\text{rd}}$ or 4$^{\text{th}}$ power. By contrast, the proposed MAD-based metrics for deviation, skewness, and kurtosis is easy to interpret and express in terms of simple differences and ratios of mean absolute deviations computed over corresponding sub-ranges of X. As a result, these measures are less impacted by outliers than the corresponding measures used in classical statistics.

Another advantage of using mean absolute deviation H instead of standard deviation σ is that H is often simpler to interpret as it is computed directly (without squaring). Consider the example suggested in [1]: if X follows a uniform distribution in $[0, 1]$ then $H = 1/4$ and $\sigma = \sqrt{3}/2$ (for details see Section 8.1). The MAD value $H = 1/4$ is easy to interpret: it represents the average distance of X from its median $M = 1/2$. However, it is more difficult to find an easy interpretation for standard deviation $\sigma = \sqrt{3}/2$.

4. MAD computation and interpretation

Our approach is to replace standard deviation σ with mean absolute deviation from the median H and to derive MAD-based measures for skewness and kurtosis without resorting to higher powers. The proposed measures are simple to interpret in terms of corresponding sub-ranges.

We will find it convenient to use the indicator function for any subset $U \subset \Omega$

$$\mathbf{1}_U = \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{otherwise} \end{cases}$$

Define the left sub-space of Ω by $\Omega^L = \{w \in \Omega \mid x(w) \leq M\}$ and right sub-space of Ω by $\Omega^R = \{w \in \Omega \mid x(w) > M\}$. Also, define $X^L = X\mathbf{1}_{\Omega^L}$ and $X^R = X\mathbf{1}_{\Omega^R}$. Then, from the definition of H in Equation (1) we have

$$H = \int_\Omega |x - M| f(x) \, dx = \int_{\Omega^L} (M - x) f(x) \, dx$$
$$+ \int_{\Omega^R} (x - M) f(x) \, dx = E(X^R) - E(X^L) \quad (2)$$

From the above equation, it is easy to show that for any constants b and c we have $H(bX + c) = |b|H$ and, therefore, H is shift invariant. In addition, from Equation (2), we can interpret H as the difference of the sub-means of the right sub-range Ω^R and left sub-range Ω^L. Since $E(X) = E(X^L) + E(X^R)$, we have $H = E(X) - 2E(X^L)$. Therefore, we can also interpret H as the difference between the mean of X and twice the sub-mean of X^L.

Let us now compare the mean absolute deviation H in Equation (2) with quartile deviation (or semi-quartile range) $H_Q = (Q_3 - Q_1)/2$ which is used in descriptive statistics as a measure of statistical dispersion [e.g., [13, 14]]: Both equations for H and H_Q have the same form under this correspondence:

$$E(X^L) \Longleftrightarrow Q_1/2 \quad \text{and} \quad E(X^R) \Longleftrightarrow Q_3/2 \quad (3)$$

Example 2: Consider the same random variable X uniform in $\Omega = \{1, 2, \ldots, 12\}$ as in the previous example. For this sequence $n = 12$, $f(x) = 1/12$, $E(X) = M = 6.5$, and $H = 3$. Let us re-compute H using the left and right half sub-ranges. To that end, we split Ω into left and right sub-spaces (Ω^L and Ω^R) around the median:

$$\Omega = \left\{ \underbrace{1, \; 2, \; 3, \; 4, \; 5, \; 6}_{\Omega^L \text{ (left half)}}, \; \underbrace{7, \; 8, \; 9, \; 10, \; 11, \; 12}_{\Omega^R \text{ (right half)}} \right\}$$

For the left sub-mean, we have $E(X\mathbf{1}_{\Omega^L}) = (1 + 2 + 3 + 4 + 5 + 6)/12 = 1.75$ and for the right sub-mean $E(X\mathbf{1}_{\Omega^R}) = (7 + 8 + 9 + 10 + 11 + 12)/12 = 4.75$. Then we compute $H = E(X^R) - E(X^L) = 3$. Alternatively, we have $H = E(X) - 2E(X^L) = 3$. To compute the corresponding quantile-based metric H_Q, we note that the first and third quartiles for this sequence are $Q_1 = 3.75$ and $Q_3 = 9.25$. Therefore, the quantile-based deviation $H_Q = (Q_3 - Q_1)/2 = 2.75$.

5. MAD-based alternatives for skewness and kurtosis

We now consider skewness and kurtosis. In classical statistics, skewness is defined as a measure of the asymmetry of the probability distribution around its mean. For a survey, see [15]. One of the most common measures is the moment coefficient of skewness (or skew) S, defined as the third standardized moment in terms of mean μ and standard deviation σ, namely,

$$S = \frac{E(X - \mu)^3}{\sigma^3} \tag{4}$$

Just like in the computation of variance, this definition is sensitive to outliers. We will define MAD-based alternatives for skewness and kurtosis without resorting to the computation of higher powers. In this way, our proposed expressions for skewness and for kurtosis will be more resilient to outliers.

We proceed as follows. From Equation (2), the expression for H can be written as follows:

$$H = E\left((M - X)\mathbf{1}_{\Omega^L}\right) \;+\; E\left((X - M)\mathbf{1}_{\Omega^R}\right) \tag{5}$$

The first term is the contribution to H from $X\mathbf{1}_{\Omega^L}$ whereas the second term is the contribution to H from $X\mathbf{1}_{\{\Omega^R}$. Our MAD-based alternative A_M for skewness can be defined as the (normalized) difference of these contributions as follows:

$$A_M = \frac{E\left((X - M)\mathbf{1}_{\Omega^R}\right) - E\left((M - X)\mathbf{1}_{\Omega^L}\right)}{H} \tag{6}$$

The above expression immediately implies that the MAD-based skewness is both shift and translation invariant.

From Equation (2), we can re-write the expression for A_M in Equation (6) as follows:

$$A_M = \frac{E(X) - M}{H} \tag{7}$$

Therefore, the MAD-based metric A_M for skewness coincides with Groeneveld and Meeden's skewness coefficient [16]. The MAD-based skewness in Equation (7) has a simple interpretation as the ratio of two distances: the numerator $E(X) - M$ is the (signed) distance between the mean and the median, whereas the numerator H is the average distance of values in X to the median. It is more difficult to attach a simple and intuitive interpretation to the classical skewness S in Equation (4).

Let us compare the MAD-based skewness from the above Equation (7) with the non-parametric skewness given by $(E(X) - M)/\sigma$. Since $H \leq \sigma$, we have the following relationship between these measures:

$$\text{Non Parametric Skewness} = \frac{E(X) - M}{\sigma} \leq A_M$$

Next, we derive the upper and lower bound for A_M. Since $E\left((X - M)\mathbf{1}_{\Omega^L}\right) < 0$ and $E\left((X - M)\mathbf{1}_{\Omega^R}\right) > 0$ from Equations (5) and (6) we easily obtain $-1 \leq A_M \leq 1$.

The definition of MAD-based skewness in Equation (6) has been suggested before [e.g., [17]]. However, our results for the representation of H would allow us to derive simple computational

expressions for MAD-skewness and compare the obtained results with skewness estimates used in quantile statistics.

To start, let us re-write our definition of A_M as follows. From Equation (6), we obtain

$$A_M = \frac{\left(E(X^R) - M/2\right) - \left(M/2 - E(X^L)\right)}{E(X^R) - E(X^L)} \tag{8}$$

We can now compare MAD-based skewness A_M from Equation (8) with the quantile skewness A_Q often used in descriptive statistics [e.g., [18, 19]]:

$$A_Q = \frac{(Q_3 - M) - (M - Q_1)}{Q_3 - Q_1} = \frac{(Q_3/2 - M/2) - (M/2 - Q_1/2)}{Q_3/2 - Q_1/2} \tag{9}$$

where Q_i denote the corresponding quartiles. The expressions in Equations (8) and (9) have the same form under the same correspondence as before in Equation (3). The numerator in Equation (9) for A_Q is the difference between the average of upper and lower quartiles and the median and the denominator is the quartile deviation H_Q. By contrast, in Equation (8) for A_M, the numerator is the difference between the concentrations of probability mass for the left and right halves and the half median, whereas the denominator is the mean absolute deviation H.

Finally, note that from Equation (9) we have

$$A_Q = -1 + \frac{2(Q_3 - M)}{Q_3 - Q_1} \geq -1 \quad \text{and} \quad A_Q = 1 - \frac{2(M - Q_1)}{Q_3 - Q_1} \leq 1$$

Therefore, both MAD-based skewness A_M and quantile-based skewness A_Q are always in the range $[-1, 1]$. By contrast, the classical skewness S can be unbounded.

We now turn our attention to kurtosis. Recall that classical Pearson's kurtosis K is defined as [20]:

$$K = \frac{E(X - \mu)^4}{\sigma^4} \tag{10}$$

To define an analogy to kurtosis using absolute deviations from the median, we find it useful to interpret kurtosis as suggested in [21]. We consider a standardized variable $Z = (X - \mu)/\sigma$ and let $W = Z^2$. Then $E(W) = 1$ and

$$\text{Var}(W) = E(W^2) - E^2(W) = K - 1 \implies K = \text{Var}(W) + 1$$

Note that since Z is normalized and dimensionless, $W = Z^2$ is also automatically dimensionless. Therefore, $\text{Var}(W)$ is also dimensionless. The kurtosis K can be viewed then as related to the dispersion of $W = Z^2$ around its mean 1. Equivalently, K is associated with the dispersion of Z around -1 and 1 [21]. This implies that kurtosis is associated with the concentration of X around points $\mu - \sigma$ and $\mu + \sigma$. High values of kurtosis can occur in a peaked unimodal distribution, in a dual-peaked bi-modal distribution, or with the concentration of probability in the tails of the distribution.

Pursuing this analogy, we want to define MAD-based kurtosis to measure the concentration of X around Q_1 (playing the role of $\mu - \sigma$) and around Q_3 (playing the role of $\mu + \sigma$) as in classical statistics.

Consider Ω^L and Ω^R defined above. We measure the concentration of probability in Ω^L around Q_1 by $E\left(\left|X - Q_1\right|\mathbf{1}_{\Omega^L}\right)$.

Similarly, we measure the concentration of probability in Ω^R around Q_3 by $\mathrm{E}\left(|X - Q_3|\mathbf{1}_{\Omega^R}\right)$. We define MAD-based alternative T_M for kurtosis by normalizing the total concentration by MAD:

$$T_M = \frac{\mathrm{E}\left(|X - Q_1|\mathbf{1}_{\Omega^L}\right) + \mathrm{E}\left(|X - Q_3|\mathbf{1}_{\Omega^R}\right)}{H} \qquad (11)$$

From the above definition, it is easy to show that MAD-based kurtosis T_M is scale and translation invariant just as MAD-based skewness A_M.

The above expression for T_M in Equation (11) can be interpreted as follows. From the definition of the median, $F(X^L) = F(X^R) = 0.5$. Therefore, the term $\mathrm{E}(|X - Q_1|\mathbf{1}_{\{X \leq M\}})$ in Equation (11) can be interpreted as one half of the average distance d_L of values of X^L from Q_1. Similarly, the term $\mathrm{E}(|X - Q_3|\mathbf{1}_{\{X > M\}})$ in Equation (11) can be interpreted as one half of the average distance d_R of values of X^R from Q_3. The numerator in Equation (11) is then the average of these distances, namely $(d_L + d_R)/2$.

Therefore, the proposed MAD-based alternative T_M for kurtosis in Equation (11) has a simple and intuitive explanation as the ratio of two distances: the numerator is the average of distances from values in Ω^L and Ω^R from Q_1 and Q_3, respectively, whereas the denominator H is the average distance of values of Ω to its median M. It is more difficult to provide an intuitive explanation of the classical kurtosis K in Equation (10).

Let us now establish some simple bounds for T_M. Recall that the median minimizes the sum of absolute deviations [11, 12]. Since Q_1 is the median for X^L and Q_3 is the median for X^R, we obtain $\mathrm{E}(|X - Q_1|\mathbf{1}_{\Omega^L}) \leq \mathrm{E}((M - X)\mathbf{1}_{\Omega^L})$ and $\mathrm{E}(|X - Q_3|\mathbf{1}_{\Omega^R}) \leq \mathrm{E}((X - M)\mathbf{1}_{\Omega^R})$. Therefore, from (11) we can immediately obtain:

$$0 \leq T_M \leq \frac{\mathrm{E}\left((M - X)\mathbf{1}_{\Omega^L}\right) + \mathrm{E}\left((X - M)\mathbf{1}_{\Omega^R}\right)}{H} = 1$$

Unlike Pearson's kurtosis K that can be unbounded [22], the proposed MAD-based alternative T_M for kurtosis is always in the range $0 \leq T_M \leq 1$. This could allow for more meaningful comparisons of data.

The definition of MAD-based measures to measure tails has been suggested before [17]. In that work, it was suggested to use $\mathrm{E}\left((M - X)\mathbf{1}_{\Omega^L}\right)/H$ and $\mathrm{E}\left((X - M)\mathbf{1}_{\Omega^R}\right)/H$ as measures of fat tails. By contrast, our suggestion for MAD-based kurtosis in Equation (11) is to use the mean absolute deviations of left and right sub-spaces from their corresponding medians Q_1 and Q_3, not from the median M. This would allow us to provide additional interpretation for the proposed kurtosis and to compare the proposed formula for T_M with quantile kurtosis T_Q suggested by Moors [21].

To proceed, let us re-write T_M in terms of sub-means of corresponding sub-means. We consider the following sub-spaces:

$$\Omega^{LL} = \{w \in \Omega \mid x(w) \leq Q_1\}, \quad \Omega^{LR} = \{w \in \Omega \mid Q_1 \leq x(w) \leq M\},$$
$$\Omega^{RL} = \{w \in \Omega \mid M \leq x(w) \leq Q_3\}, \quad \Omega^{RR} = \{w \in \Omega \mid x(w) \geq Q_3\}$$

and define

$$X^{LL} = X\mathbf{1}_{\Omega^{LL}}, \quad X^{LR} = X\mathbf{1}_{\Omega^{LR}},$$
$$X^{RL} = X\mathbf{1}_{\Omega^{RL}}, \quad X^{RR} = X\mathbf{1}_{\Omega^{RR}}$$

From the above definitions, we have

$$\mathrm{E}(|X - Q_1|\mathbf{1}_{\{X \leq M\}}) = \mathrm{E}(X^{LR}) - \mathrm{E}(X^{LL}),$$
$$\mathrm{E}(|X - Q_3|\mathbf{1}_{\{X > M\}}) = \mathrm{E}(X^{RR}) - \mathrm{E}(X^{RL})$$

and therefore, we can re-write our expression (11) for kurtosis T_M as follows

$$T_M = \frac{\mathrm{E}(X^{LR}) - \mathrm{E}(X^{LL}) + \mathrm{E}(X^{RR}) - \mathrm{E}(X^{RL})}{\mathrm{E}(X^R) - \mathrm{E}(X^L)} \qquad (12)$$

Moors [21] suggested a quantile-based formula for kurtosis in terms of the octiles $O_1, \ldots, O_7$ as follows:

$$T_Q = \frac{(O_7 - O_5) + (O_3 - O_1)}{(O_6 - O_2)} \qquad (13)$$

The expressions in Equations (12) and (13) have the same form under the following correspondence:

$$\mathrm{E}(X^{LR}) - \mathrm{E}(X^{LL}) \iff \frac{(O_3 - O_1)}{2},$$
$$\mathrm{E}(X^{RR}) - \mathrm{E}(X^{RL}) \iff \frac{(O_7 - O_5)}{2},$$
$$\mathrm{E}(X^R) - \mathrm{E}(X^L) \iff \frac{(O_6 - O_2)}{2}$$

Our justification for (12) is analogous to the justification for a quantile-based alternative to kurtosis in (13) in terms of octiles suggested in [21]. The terms in the numerator of T_Q are large if large probability mass is concentrated in O_2 and O_6 corresponding to large dispersion around $\mu - \sigma$ and $\mu + \sigma$. The terms in the numerator of T_M are small if small probability mass is concentrated in Q_1 and Q_3. The difference between our formula in (12) and the quantile-based formula in (13) is that we measure these masses in terms of "partial" means $\mathrm{E}(X^{LL})$, $\mathrm{E}(X^{LR})$, $\mathrm{E}(X^{RL})$ and $\mathrm{E}(X^{RR})$ instead of octiles. This is illustrated in the following example.

Example 3: As before, consider a random variable X with uniform probability in $\Omega = \{1, 2, \ldots, 12\}$. The corresponding sub-ranges are shown below:

$$\Omega = \underbrace{\underbrace{1, \ 2, \ 3}_{\Omega^{LL}\ (1^{st}\ \text{quarter})}, \ \underbrace{4, \ 5, \ 6}_{\Omega^{LR}\ (2^{nd}\ \text{quarter})}}_{\Omega^L\ (\text{left half})}, \ \underbrace{\underbrace{7, \ 8, \ 9}_{\Omega^{RL}\ (3^{rd}\ \text{quarter})}, \ \underbrace{10, \ 11, \ 12}_{\Omega^{RR}\ (4^{th}\ \text{quarter})}}_{\Omega^R\ (\text{right half})}$$

For this sequence, we compute $H = 3$. This distribution is symmetric and $S = A_M = A_Q = 0$. We will therefore focus on MAD-based kurtosis T_M. The median for the left sub-range Ω^L is $Q_1 = 3.5$ and we compute $H(X^L, Q_1) = 0.75$. The value $H(X^L, Q_1) = 1.5$ is one half of the average distance d_L from values in Ω^L to its median Q_1. Similarly, the median for the right sub-range Ω^R is $Q_3 = 9.5$ and we compute $H(X^R, Q_3) = 0.75$. The value $H(X^R, Q_3) = 1.5$ is one half of the average distance d_R from values in Ω^R to its median Q_R. Therefore, the term $H(X^L, Q_1) + H(X^R, Q_3) = (d_L + d_R)/2 = 0.75$ is the average of the distances from sub-ranges Ω^L and Ω^R to their respective median. From Equation (11) the MAD-based kurtosis $T_M = 0.5$. It has a simple

interpretation as the ratio of the average distances $(d_L + d_R)/2$ to H which is the average distance of Ω to its median $M = 6.5$. Let us re-compute T_M using sub-means. We have $E(X^{LL}) = 0.5$, $E(X^{LR}) = 1.25$, $E(X^{RL}) = 2$ and $E(X^{RR}) = 2.75$. Applying Equation (12) we have $T_M = 0.5$. By contrast, the standard statistical kurtosis is $K = 1.78$. Let us compare the MAD-based kurtosis T_M with quantile-based kurtosis T_Q in formula (13). To compute T_Q, we compute the octiles $O_1, \ldots, O_7$ for our sequence (using midpoint for interpolation): $O_1 = 2.375$ $O_2 = 3.75$, $O_3 = 5.125$, $O_4 = 6.5$, $O_5 = 7.875$, $O_6 = 9.25$ and $O_7 = 10.625$. Then, the quantile formula for kurtosis from Equation (13) gives us $T_Q = 1$.

6. Discussion

One of the reasons to use quantile-based metrics for data is that they are resistant to outliers. One of the measures of this resistance is the sample breakdown point - the proportion of observations that can be altered that can result in statistics being arbitrarily large or small. The median M has a breakpoint of 50%. This means that 50% of the points must be "outliers" before the median can be moved outside the range of outliers [9, 23]. By contrast, the breakpoint of the mean is 0%. This means that a single observation would change it.

The computation of MAD-based measures for deviation, skewness, and kurtosis involves computing the corresponding sub-spaces. A single change in observation value would change these measures. However, these changes will not be as dramatic as changes for the classical measures.

By contrast, the MAD-based measures are determined by the corresponding sub-means. Any change in values would change some of these, and this, in turn, would result in different values for MAD-based measures for each of the sub-spaces. Note that a single observation value change would change the corresponding partial mean without affecting other sub-means. By contrast, a single change in observation would change the mean and result in changes for the corresponding moments, affecting both skewness and kurtosis. However, because the proposed MAD-based measures involve only the first moment whereas classical skewness and kurtosis involve the 3rd and 4th moments, respectively, we would expect that MAD-based measures would change by a smaller percentage than the corresponding classical measures.

For illustration, consider the MAD-based alternative T_M for kurtosis. It can capture concentrations in the tails more accurately than can the quantile kurtosis T_Q. For example, if the largest 10% of values in X increase in value then this cannot be captured by T_Q since octiles will not change. Similarly, if the smallest 10% of values in X decrease, then again octiles will not change resulting in the same value for octile kurtosis. By contrast, the proposed MAD-based formula for kurtosis uses sub-means of appropriate sub-ranges and can, therefore, more accurately reflect the impact of such larger or smaller values.

Schematically, we can indicate this as follows:

$$\Omega = \underbrace{\overbrace{\ldots O_1 \ldots}^{\Omega^{LL}\ (1^{st}\ \text{quarter})} Q_1 \overbrace{\ldots O_3 \ldots}^{\Omega^{LR}\ (2^{nd}\ \text{quarter})}}_{\Omega^L\ (\text{left half})},$$

$$M, \underbrace{\overbrace{\ldots O_5 \ldots}^{\Omega^{RL}\ (3^{rd}\ \text{quarter})} Q_3 \overbrace{\ldots O_7 \ldots}^{\Omega^{RR}\ (4^{th}\ \text{quarter})}}_{\Omega^R\ (\text{right half})}$$

If we consider any changes in values in quarter sub-ranges Ω^{LL}, Ω^{LR}, Ω^{RL} or Ω^{RR} that do not change the octiles, then the quantile kurtosis T_Q would remain the same. By contrast, these changes in values will change the corresponding sub-means and, therefore the value of MAD-based kurtosis T_M. Therefore, MAD-based kurtosis T_M can capture changes in probability mass in the tails more accurately than using octiles in T_Q.

By the same argument, it is easy to show that quantile skewness A_Q and quantile deviations H_Q would remain unchanged, whereas the mean absolute deviation from median H and MAD-based skewness A_M would change.

The proposed MAD-based alternative measures for deviation, skewness, and kurtosis provide additional tools for data analysis. These measures are less sensitive to outliers in that they change by a smaller percentage as compared to the changes in the classical statistics metrics.

At the same time, they do not ignore outliers as quantile-based measures do. In situations where classical, MAD-based, and quantile-based kurtosis could be computed, MAD-based kurtosis T_M has the advantage of $0 \leq T_M \leq 1$, whereas both the classical kurtosis S could be unbounded. For some distributions such as log-normal, the quantile-based kurtosis could be unbounded as well (see Section 8.3). Using the proposed MAD-based alternative measure for kurtosis with a value that is always in $[0, 1]$ provides a potentially useful tool to directly compare distributions in terms of the concentration of data in the tails.

As one potentially important application of these proposed measures considers the national income and wealth distributions [24]. It is widely recognized that there are disproportionate concentrations of income and wealth at the highest quantile. As the above example illustrates, the MAD-based metrics appear to have the ability to characterize concentrations in the upper-most quantiles in a manner that is not possible with the classical and quantile-based methods. This ability likely follows from the property that the MAD-based metrics are sensitive to excessive concentrations in the upper-most quantile, while the quantile-based metrics are not. At the same time, the classical metrics that use the third and fourth moment may overly exaggerate the impact of these concentrations in the upper-most quantile. The foregoing analysis demonstrates that these alternative MAD-based metrics would be able to capture the detailed behavior that is engendered by the concentrations of extreme income and wealth at the highest range of distribution, namely the highest quantile. We illustrate this by a detailed example in Section 9.

7. Computational considerations

In the computation of quantile-based measures, we need to compute the quantiles. These can be obtained from the inverse of the cumulative distribution function $F^{-1}(p)$. The quartiles Q_1, M and Q_3 are obtained as $Q_1 = F^{-1}(1/4)$, $M = F^{-1}(1/2)$ and $Q_3 = F^{-1}(3/4)$ whereas the remaining octiles are $O_i = F^{-1}(i/8)$ for $i = 1, 3, 5, 7$. Then the quantile-based measures for deviation, skewness, and kurtosis are

$$H_Q = \frac{Q_3 - Q_1}{2}, \quad A_Q = \frac{Q_3 + Q_1 - 2M}{(Q_3 - Q_1)},$$
$$T_Q = \frac{(O_7 - O_5) + (O_3 - O_1)}{(O_6 - O_2)} \tag{14}$$

If the distribution is symmetric, the skewness $A_Q = 0$ and for octiles we have $O_1 = 2M - O_7$, $Q_1 = 2M - Q_3$, and $O_3 = 2M - O_5$. In particular, $(O_7 - O_5) = (O_3 - O_1)$. In this case, we only need to compute four quantiles, namely M, O_5, Q_3 and Q_7 to obtain

$$H_Q = Q_3 - M \quad \text{and} \quad T_Q = \frac{2(O_7 - O_5)}{(Q_3 - M)} \tag{15}$$

Note that the quantile-based measures are expressed in terms of differences of corresponding quantiles. This implies, in particular, that the above quantile-based measures in Equation (14) are shift-invariant.

In the computation of MAD-based performance measures, we need to compute the corresponding sub-means. To facilitate this computation, consider the following auxiliary integral:

$$I(z) = \int_{t \leq z} t f(t) \, dt$$

If we can evaluate the above integral for $I(Q_1)$, $I(M)$, and $I(Q_3)$ then we can compute MAD-based performance measures from these integrals and the mean value $E(X)$. Specifically, $E(X^L) = I(M)$ and $E(X^R) = E(X) - I(M)$. For the other sub-means, we have $E(X^{LL}) = I(Q_1)$, $E(X^{LR}) = I(M) - I(Q_1)$, $E(X^{RL}) = I(Q_3) - I(M)$, and $E(X^{RR}) = E(X) - I(Q_3)$. From this, we obtain the following for MAD-based deviation, skewness and kurtosis:

$$H = E(X) - 2I(M), \quad A_M = \frac{E(X) - M}{E(X) - 2I(M)},$$
$$T_M = \frac{E(X) + 2I(M) - 2I(Q_1) - 2I(Q_3)}{E(X) - 2I(M)} \tag{16}$$

In some situations, it is easier to evaluate the following integral

$$J(z) = \int_{t \geq z} t f(t) \, dt$$

Since $J(z) = E(X) - I(z)$, we can compute MAD-based deviation, skewness, and kurtosis as follows:

$$H = 2J(M) - E(X), \quad A_M = \frac{E(X) - M}{2J(M) - E(X)},$$
$$T_M = \frac{2J(Q_1) + 2J(Q_3) - 2J(M) - E(X)}{2J(M) - E(X)} \tag{17}$$

The computation of MAD-based performance measures requires the computation of $I(Q_1)$, $I(M)$ and $I(Q_3)$ in addition

to computing the expected value $E(X)$. By contrast, in classical statistics, to compute skewness, we need to compute expectation $E(X)$, standard deviation σ, and the third moment $E(X^3)$. To compute Pearson's kurtosis would require us also to compute the fourth moment $E(X^4)$. Therefore, the computational cost of computing these measures is not higher than that of computing Pearson measures of classical statistics. Moreover, unlike the classical measures, the computation of MAD-based measures requires only the existence of first-order moments. For example, consider Pareto distributions with parameter α (see Section 8.6). For such distributions, the mean is defined for $\alpha > 1$, variance is defined for $\alpha > 2$, and kurtosis is defined for $\alpha > 3$. By contrast, MAD-based measures would require only $\alpha > 1$. Therefore, for $1 < \alpha < 2$ we can only use MAD-based or quantile-based measures to analyze deviation, skewness, and kurtosis.

On the other hand, we should note that there are situations where classical kurtosis K or MAD kurtosis T_M does not exists but the quantile-based kurtosis T_Q exists and is finite. An example is presented in [21]. If X follows the Cauchy distribution, its expected value, variance, and kurtosis are undefined. However, the quantile-based kurtosis T_Q is finite with $T_Q = 2$.

In most situations where classical, MAD-based, and quantile-based kurtosis could be computed, MAD-based kurtosis T_M has the advantage of $0 \leq T_M \leq 1$ whereas the classical kurtosis K can be unbounded. For some distributions such as log-normal, the quantile-based kurtosis T_Q could also be unbounded (see Section 8.3 below for details). Using a MAD-based measure T_M for kurtosis with a value in $0 \leq T_M \leq 1$ allows an additional comparison of all distributions in terms of their tails.

8. Comparisons for distributions

We now turn our attention to some well-known distributions. We will compute MAD-based alternatives for deviation, skewness, and kurtosis and compare them with the corresponding quantile-based and classical metrics for the following well-known probability distributions: continuous uniform, normal, log-normal, exponential, Laplace, and Pareto distributions. A summary table is provided in Section 1.

8.1. Continuous uniform distribution

Suppose X is distributed according to a uniform distribution in $[a, b]$. Its density $f(x) = 1/(b - a)$ and its cumulative distribution function $F(x) = (x - a)/(b - a)$. For this distribution, $E(X) = M = (a + b)/2$, $\sigma = (b - a)/2\sqrt{3}$ and $K = 9/5$. Since this distribution is symmetric, the skewness measures are 0.

The quantiles are computed from from $F^{-1}(p) = (1 - p)a + pb$. In particular, $Q_1 = (3a + b)/4$, $M = (a + b)/2$ and $Q_3 = (a + 3b)/4$. Since this distribution is symmetric, we compute the octiles $O_1 = (7a + b)/8$, $O_3 = (5a + 3b)/6$, $O_5 = (3a + 5b)/8$ and $O_7 = (a + 7b)/8$. To compute MAD-based measures, consider the following integral

$$J(z) = \int_a^z t f(t) \, dt = \frac{z^2 - a^2}{2(b - a)}, \qquad a \leq z \leq b$$

TABLE 1 A comparison of measures for uniform distribution.

Measure	Classical	MAD-based	Quantile-based
Deviation $H < \sigma < H_Q$	$\sigma = \dfrac{(b-a)}{2\sqrt{3}}$	$H = \dfrac{(b-a)}{4}$	$H_Q = \dfrac{(b-a)}{2}$
Skewness	$S = 0$	$A_M = 0$	$A_Q = 0$
Kurtosis $T_M < T_Q < K$	$K = 1.8$	$T_M = 0.5$	$T_Q = 1$

We have $J(Q_1) = (Q_1^2 - a^2/2(b-a))$, $J(M) = (M^2 - a^2)/2(b-a)$ and $J(Q_3) = (Q_3^2 - a^2)/2(b-a)$. From this using Equation (17) and Equation (15) we compute MAD-based and quantile-based measures for deviation and kurtosis. The results are summarized in the Table 1.

8.2. Gaussian distribution

Suppose X is distributed according to normal distribution $N(\mu, \sigma^2)$ with density $f(x)$ and cumulative distribution function $F(x)$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2},$$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2}\, dt$$

where $\Phi(\cdot)$ denotes the cumulative distribution of the standard normal. Let Q_1 and Q_3 denote the first and third quartiles of the standard normal distribution. This distribution is symmetric; therefore, all skewness measures are 0.

The MAD-based, quantile-based, and classical measures are invariant under shifts. Skewness and kurtosis are also invariant under scaling whereas for MAD-based deviation, $H(X/\sigma) = (1/\sigma)H(X)$. Therefore, we can consider the standard normal distribution for X and multiply the obtained value for H(X) by $1/\sigma$.

The quartiles for this distribution $Q_3 = -Q_1 \approx 0.67$ whereas for the octiles we have $O_5 = -O_3 \approx 0.32$, $Q_3 = -Q_1 \approx 0.68$ and $O_7 = -O_1 \approx 0.15$. To compute MAD-based MAD-based measures, consider the following integral

$$J(z) = \int_{z}^{\infty} tf(t)\, dt = \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} te^{-t^2/2}\, dt = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Since $Q_3 = -Q_1 \approx 0.67$ we have $J(Q_1) = J(Q_3) = \exp(-Q_3^2/2)/\sqrt{2\pi}$ and $J(0) = 1/\sqrt{2\pi}$. Therefore, from the above, using Equation (17) and Equation (15) we compute MAD-based and quantile-based measures for deviation and kurtosis. The results are summarized in Table 2.

8.3. Log-normal distribution

Suppose X is distributed according to log-normal distribution with parameters $\mu \in (\infty, +\infty)$ and σ^2 ($\sigma > 0$). Its density $f(x)$

and its cumulative distribution function $F(x)$ are given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\log x - \mu)^2/2\sigma^2}, \quad \text{and} \quad F(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal.

We will use the apostrophe $'$ to distinguish the performance measures of X in the log-normal distribution from those of the underlying normal distribution. Therefore, σ' will denote the standard deviation of X, μ' will denote the mean of X etc. As before, let O_i denote the octiles of the standard normal and let O_i' denote the octiles of the log-normal distribution. Then $O_i' = e^{\mu + \sigma O_i}$. In particular, the log-normal median is $M' = e^{\mu}$ and the log-normal mean is $\mu' = e^{\mu + \sigma^2/2}$. Similarly, let Q_1 and Q_3 denote the first and third quartiles for the standard normal distribution and let Q_1' and Q_3' denote the corresponding quartiles for the log-normal distribution. Consider the following integral (derived in Appendix A2)

$$I(z) = \int_{0}^{z} tf(t)\, dt = \mu' \Phi\left(\frac{\log z - \mu}{\sigma} - \sigma\right)$$

In particular, $I(Q_1') = \mu'(1 - \Phi(\sigma - Q_1))$, $I(M') = \mu'(1 - \Phi(\sigma))$ and $I(Q_3') = \mu'(1 - \Phi(\sigma - Q_3))$. Therefore, from the above, using Equation (17) and Equation (15) we compute MAD-based and quantile-based measures for deviation and kurtosis. Note that we can express mean absolute deviation H' in terms of the error function erf($\cdot$). Using $\text{erf}(x) = 2\Phi(x\sqrt{2}) - 1$ we obtain $H' = 2\mu'\,\text{erf}(\sigma/\sqrt{2})$.

In Appendix A2, we show the following relationships between MAD-based and quantile-based measures:

$$H_Q' < H' \leq \sigma', \quad A_M', A_Q' < S', \quad T_M' < T_Q' < K'$$

We summarize our results in Table 3. In addition, in Appendix A2, we examined the asymptotic behavior of these measures for $\sigma \mapsto 0$ and for $\sigma \mapsto \infty$. For $\sigma \mapsto 0$ we have $A_M' \mapsto 0$, $T_M' \mapsto 0.59$, $A_Q' \mapsto 0$ and $T_Q' \mapsto 1.23$. Therefore, for $\sigma \mapsto 0$, the MAD-based and quantile-based measures for skewness and kurtosis converge to the corresponding values for normal distribution. On the other hand, for $\sigma \mapsto \infty$

we have: $A_M' \mapsto 1$, $T_M' \mapsto 1$, $A_Q' \mapsto 1$ and $T_Q' \mapsto \infty$. Finally, note that for log-normal distribution both quantile kurtosis T_Q' and classical kurtosis K' are unbounded whereas the MAD-based kurtosis T_M' always satisfies $0 \leq T_M' \leq 1$. This allows us to compare distribution in the tails across different distributions. The results are summarized in Table 3.

8.4. Exponential distribution

Suppose X is distributed according to an exponential distribution with rate $\lambda > 0$ [22]. Its density $f(x) = \lambda e^{-\lambda x}$ and its cumulative distribution function $F(x) = 1 - e^{-\lambda x}$ with $x \in [0, \infty)$. Its mean $E(X) = 1/\lambda$ and its standard deviation $\sigma = 1/\lambda$. The quantiles of exponential distribution are $F^{-1}(p)$ and are given by $-\log(1-p)/\lambda$. In particular, $Q_1 = \log(4/3)/\lambda$, $M = \log(2)/\lambda$

TABLE 2 A comparison of measures for normal distribution.

Measure	Classical	MAD-based	Quantile-based
Deviation $H_Q < H < \sigma$	σ	$H = \sqrt{\dfrac{2}{\pi}}\,\sigma \approx 0.80\,\sigma$	$H_Q = Q_3\,\sigma \approx 0.68\,\sigma$
Skewness	$S = 0$	$A_M = 0$	$A_Q = 0$
Kurtosis $T_M < T_Q < K$	$K = 3$	$T_M = -1 + 2e^{-Q_3^2/2} \approx 0.59$	$T_Q = \dfrac{O_7 - O_5}{Q_3} \approx 1.23$

TABLE 3 A comparison of measures for log-normal distribution.

Measure	Classical	MAD-based	Quantile-based
Deviation $H'_Q < H' < \sigma'$	$\sigma' = \sqrt{(e^{\sigma^2}-1)e^{2\mu+\sigma^2}}$	$H' = e^{\mu+\sigma^2/2}\left(2\Phi(\sigma)-1\right)$	$H'_Q = e^{\mu}\,\dfrac{(e^{\sigma Q_3}-e^{-\sigma Q_3})}{2}$
Skewness $0 < A'_M, A'_Q < S'$	$S' = (e^{\sigma^2}+2)\sqrt{e^{\sigma^2}-1}$	$A'_M = \dfrac{(e^{\sigma^2/2}-1)}{e^{\sigma^2/2}(2\Phi(\sigma)-1)}$	$A'_Q = -1 + \dfrac{2e^{\sigma Q_3}}{(e^{\sigma Q_3}+1)}$
Kurtosis $T'_M < T'_Q < K'$	$K' = e^{4\sigma^2} + 2e^{3\sigma^2}$ $+3e^{2\sigma^2} - 6$ (unbounded)	$T'_M = \dfrac{N_M(\sigma,Q_1,Q_3)}{2\Phi(\sigma)-1}$ (bounded) $0 < T_M < 1$	$T'_Q = \dfrac{N_Q(\sigma,O_5,O_7)}{(e^{\sigma Q_3}-e^{-\sigma Q_3})}$ (unbounded)
	$N_M(\sigma, Q_1, Q_3) = -1 - 2\Phi(\sigma) + 2\Phi(\sigma - Q_1) + 2\Phi(\sigma - Q_3)$		
	$N_Q(\sigma, O_5, O_7) = (e^{-\sigma O_5} - e^{-\sigma O_7}) + (e^{\sigma O_7} - e^{\sigma O_5})$		

and $Q_3 = \log(4)/\lambda$ whereas the octiles are $O_1 = \log(8/7)/\lambda$, $O_3 = \log(8/5)/\lambda$, $O_5 = \log(8/3)/\lambda$ and $O_7 = \log(8)/\lambda$.

To compute MAD-based measures, we compute (using integration by parts) the following integral

$$I(z) = \int_0^z t f(t)\, dt = \int_0^z \lambda t e^{-\lambda t}\, dt = \frac{1 - (1 + \lambda z)e^{-\lambda z}}{\lambda}$$

We compute $I(Q_1) = (1 - 3\log(4/3))/4\lambda$, $I(M) = (1 - \log 2)/2\lambda$ and $I(Q_3) = (3 - \log 4)/4\lambda$. Then from the above results and from Equations (16 a)nd (15) we compute MAD-based and quantile-based measures for deviation, skewness, and kurtosis. The results are summarized in the Table 4.

8.5. Laplace distribution

Suppose X is distributed according to Laplace distribution with location μ and scale b. Its density $f(x)$ and cumulative distribution function $F(x)$ are given by Feller [22]

$$f(x) = \frac{1}{2b}e^{-|x-\mu|/b} \quad \text{and} \quad F(x) = \begin{cases} \frac{1}{2}e^{(x-\mu)/b}, & \text{if } x \leq \mu \\ 1 - \frac{1}{2}e^{-(x-\mu)/b}, & \text{if } x \geq \mu \end{cases}$$

This distribution of X is symmetric around μ. its median M and mean $E(X)$ are both the same with $E(X) = M = \mu$. Its standard deviation $\sigma = b\sqrt{2}$. Since both MAD-based measures and quantile-based measures are shift-invariant, we let $\mu = 0$.

The quantiles are easily computed as $F^{-1}(p)$. In particular, $Q_1 = -b\log 2$ and $Q_3 = b\log 2$. whereas for the octiles we have $O_1 = -2b\log 2$, $O_3 = b\log 3 - 2b\log 2$, $O_5 = 2b\log 2 - b\log 3$, and $O_7 = 2b\log 2$.

To compute the MAD-based measures, we compute the integral $J(z)$ (derived in Appendix A3):

$$J(z) = \int_z^\infty x f(x)\, dx = \begin{cases} \dfrac{(b+z)}{2}\, e^{-z/b}, & z \geq 0 \\ \dfrac{(b-z)}{2}\, e^{z/b}, & z < 0 \end{cases}$$

We compute $J(Q_1 = b(1 + \log 2)/4$, $J(M) = b/2$ and $J(Q_3) = b(1 + \log 2)/4$. Then from Equations (17) and (15) we can compute MAD-based and quantile-based measures for deviation, skewness, and kurtosis. We summarize our results in Table 5.

We note that just as for classical kurtosis, the MAD-based kurtosis for Laplace distribution is greater than the MAD-based kurtosis for Normal distribution. This is because the Laplace distribution has fatter tails compared to the Normal distribution.

8.6. Pareto distribution

Suppose X is distributed according to Pareto Type I distribution distribution with shape $\alpha > 0$ and scale $\beta > 0$. Its density $f(x)$ and its cumulative distribution function $F(x)$ are given by

$$f(x) = \begin{cases} \dfrac{\alpha\beta^\alpha}{x^{\alpha+1}} & x \geq \beta \\ 0 & x < \beta \end{cases} \quad \text{and} \quad F(x) = \begin{cases} 1 - \left(\dfrac{\beta}{x}\right)^\alpha & x \geq \beta \\ 0 & x < \beta \end{cases}$$

This distribution has infinite mean μ for $\alpha \leq 1$, undefined variance σ^2 for $\alpha \leq 2$, undefined skewness S for $\alpha \leq 3$ and undefined (excess) kurtosis K for $\alpha \leq 4$. The quantiles are computed from $F^{-1}(p)$ and are given by $\beta/(1-p)^{1/\alpha}$. For Q_1, M

TABLE 4 A comparison of measures for exponential distribution.

Measure	Classical	MAD-based	Quantile-based
Deviation $H < H_Q < \sigma$	$\sigma = \dfrac{1}{\lambda}$	$H = \dfrac{\log 2}{\lambda} \approx 0.69\sigma$	$H_Q = \dfrac{\log 3}{2\lambda} \approx 0.55\sigma$
Skewness $A_Q < A_M < S$	$S = 2$	$A_M = -1 + \dfrac{1}{\log 2} \approx 0.44$	$A_Q = -1 + \dfrac{2\log 2}{\log 3} \approx 0.26$
Kurtosis $T_M < T_Q < K$	$K = 9$	$T_M = 3 - \dfrac{3\log 3}{2\log 2} \approx 0.62$	$T_Q = 1 + \dfrac{\log(7/5)}{\log 3} \approx 1.31$

TABLE 5 A comparison of measures for Laplace distribution.

Measure	Classical	MAD-based	Quantile-based
Deviation $H_Q < H < \sigma$	$\sigma = b\sqrt{2} \approx 1.41b$	$H = b$	$H_Q = b\log 2 \approx 0.69b$
Skewness	$S = 0$	$A_M = 0$	$A_Q = 0$
Kurtosis $T_M < T_Q < K$	$K = 6$	$T_M = \log 2 \approx 0.69$	$T_Q = \dfrac{\log 3}{\log 2} \approx 1.58$

TABLE 6 A comparison of measures for Pareto distribution.

Measure	Classical	MAD-based	Quantile-based
Deviation $H_Q < H < \sigma$	$\sigma = \dfrac{\beta}{(\alpha-1)}\sqrt{\dfrac{\alpha}{\alpha-2}}$ $(\alpha > 2)$	$H = \dfrac{\beta\alpha(\sqrt[\alpha]{2}-1)}{(\alpha-1)}$ $(\alpha > 1)$	$H_Q = \dfrac{\beta(\sqrt[\alpha]{4}-\sqrt[\alpha]{4/3})}{2}$
Skewness $A_Q < A_M < S$	$S = \dfrac{2(1+\alpha)}{(\alpha-3)}\sqrt{\dfrac{\alpha-2}{\alpha}}$ $(\alpha > 3)$	$A_M = -1 + \dfrac{\sqrt[\alpha]{2}}{\alpha(\sqrt[\alpha]{2}-1)}$	$A_Q = -1 + \dfrac{2(\sqrt[\alpha]{4}-\sqrt[\alpha]{2})}{\sqrt[\alpha]{4}-\sqrt[\alpha]{4/3}}$
Kurtosis $T_M < T_Q < K$	$K = \dfrac{6(\alpha^3+\alpha^2-6\alpha-2)}{\alpha(\alpha-3)(\alpha-4)}$ $(\alpha > 4)$	$T_M = \dfrac{N_M(\alpha)}{2(\sqrt[\alpha]{2}-1)}$	$T_Q = \dfrac{N_Q(\alpha)}{\sqrt[\alpha]{4}-\sqrt[\alpha]{4/3}}$
	$N_M(\alpha) = 3\sqrt[\alpha]{4/3} + \sqrt[\alpha]{4} - 2\sqrt[\alpha]{2} - 2$		
	$N_Q(\alpha) = (\sqrt[\alpha]{8/5} - \sqrt[\alpha]{8/7}) + (\sqrt[\alpha]{8} - \sqrt[\alpha]{8/3})$		

and Q_3 we have $Q_1 = \beta\sqrt[\alpha]{4/3}$, $M = \beta\sqrt[\alpha]{2}$ and $Q_3 = \beta\sqrt[\alpha]{4}$. For the octiles we have $O_1 = \sqrt[\alpha]{8/7}$, $O_3 = \sqrt[\alpha]{8/5}$, $O_5 = \sqrt[\alpha]{8/3}$ and $O_7 = \sqrt[\alpha]{8}$. To compute MAD-based measures, we compute the integral for any $z >= \beta$ and $\alpha > 1$

$$J(z) = \int_z^\infty xf(x)\,dx = -\frac{\alpha\beta^\alpha}{(1-\alpha)}x^{-\alpha+1}|_z^\infty = \frac{\alpha z}{(\alpha-1)}\left(\frac{\beta}{z}\right)^\alpha$$
$$= \frac{\alpha z(1-F(z))}{(\alpha-1)}$$

Since $1 - F(Q_1) = 3/4$, $1 - F(M) = 1/2$, and $1 - F(Q_3) = 1/4$, we immediately compute $J(Q_1) = 3\alpha\beta Q_1/4(\alpha-1)$, $J(M) = \alpha\beta M/2(\alpha-1)$, and $J(Q_3) = \alpha\beta Q_3/3(\alpha-1)$. Then from Equations (17) and (15) we can compute MAD-based and quantile-based measures for deviation H_Q, skewness A_Q, and kurtosis T_Q. Moreover, in Appendix (Section A4), we showed the following relationship between measures: $H_Q < H < \sigma$ and $A_Q < A_M < S$ and $T_M < T_Q < K$. These results are summarized in Table 6.

In Appendix (Section A4), we examined the asymptotic behavior of these measures for $\alpha \mapsto 1$ and for $\alpha \mapsto \infty$. For $\alpha \mapsto 1$,

we showed that for MAD-based measures that $H \mapsto \infty$, $A_M \mapsto 1$, and $T_M \mapsto 1$ whereas for quantile-based measures $H_Q \mapsto 4/3$, $A_Q \mapsto 0.5$, and $T_Q \mapsto 2.17$. By contrast, for $\alpha \mapsto \infty$ we showed in Appendix (Section A4) that both MAD-based and quantile-based measures for skewness and kurtosis converge to the corresponding measures for exponential distribution $H \mapsto 0$, $A_M \mapsto 0.44$, $T_M \mapsto 0.62$, $H_Q \mapsto 0$, $A_Q \mapsto 0.26$, $T_Q \mapsto 1.31$ (see Table 4). Note that if X is Pareto with shape α and scale β, then $Y = \log(X/\beta)$ is exponentially distributed with rate α [20, 22].

9. Example: wealth distribution

Assume that income is distributed according to a Pareto principle $p + q$ principle [20]: $100p\%$ of all income is received by $100q\%$ of people $(p + q = 1)$ For example, the $60 - 40$ rule $(p = 0.6, q = 0.4)$ means that 40% of the people receive 60% of the wealth. We assume that $p > 0.5$.
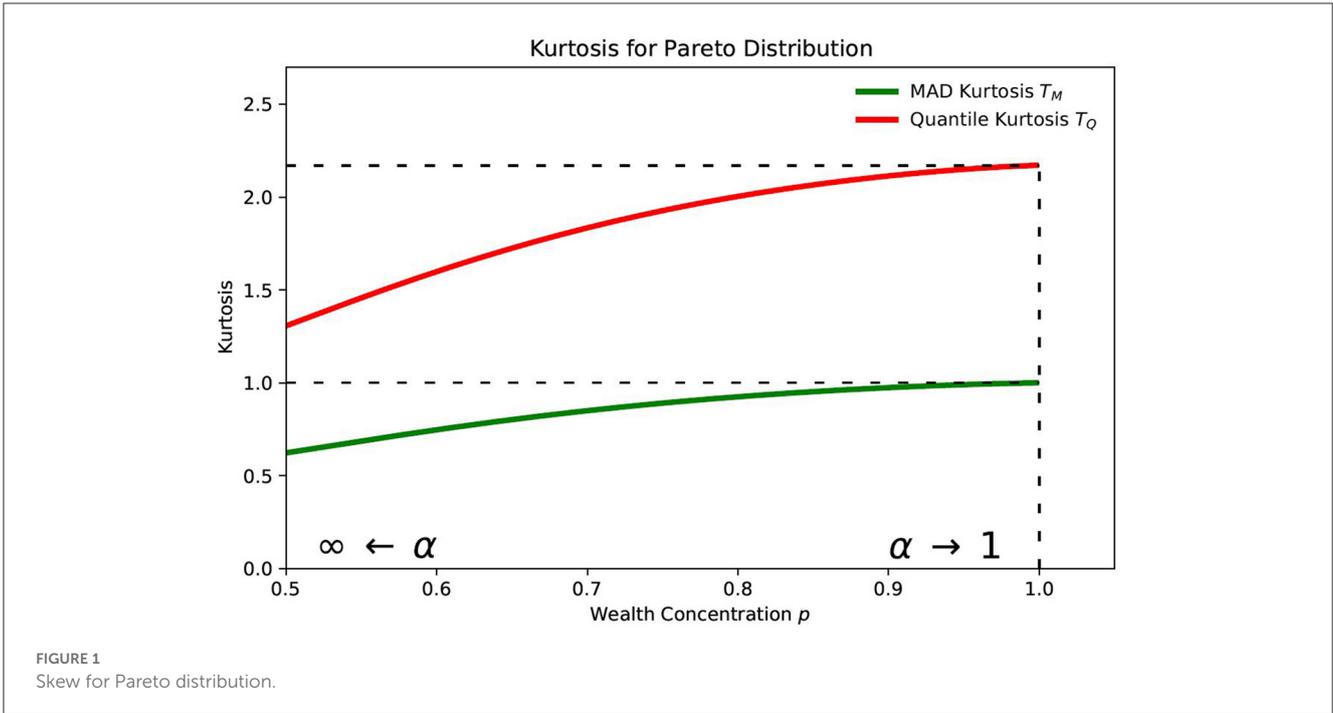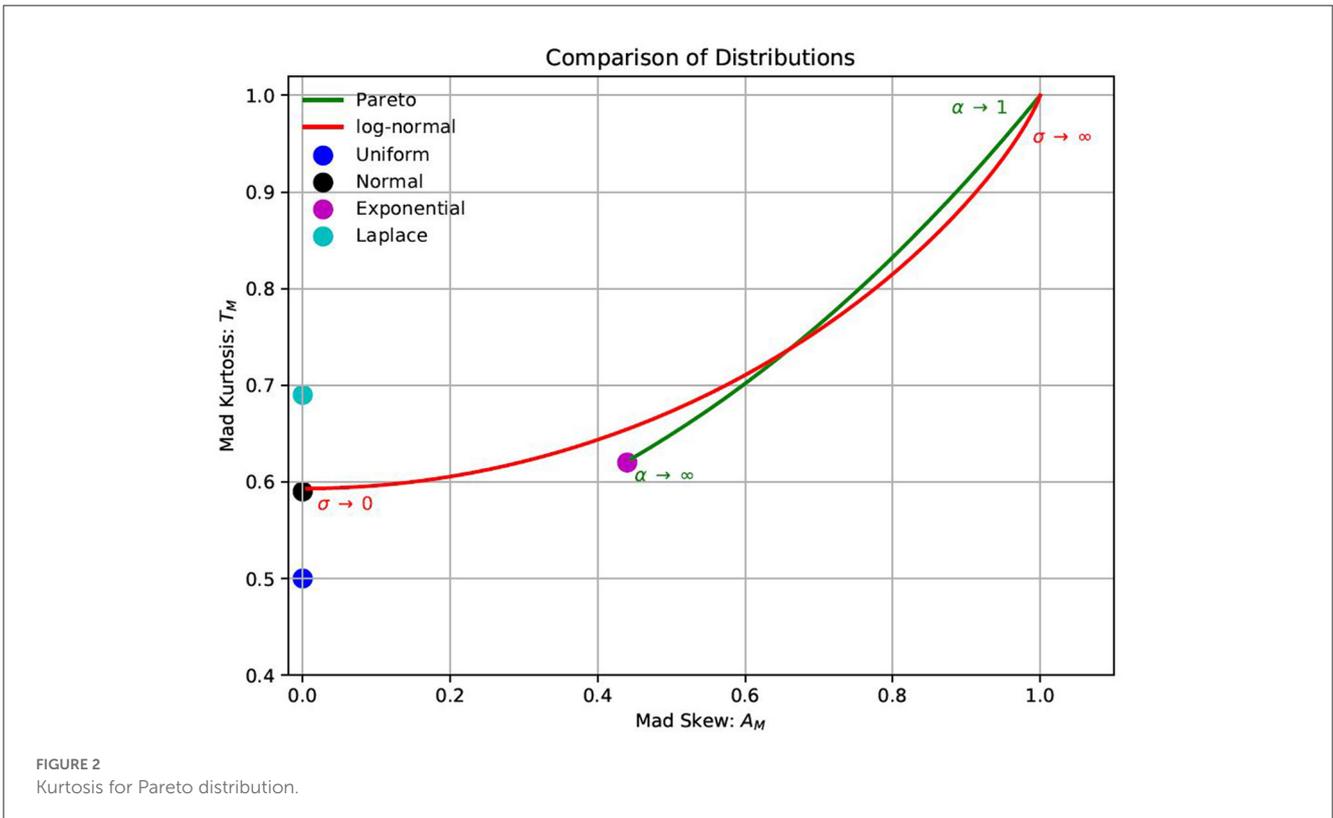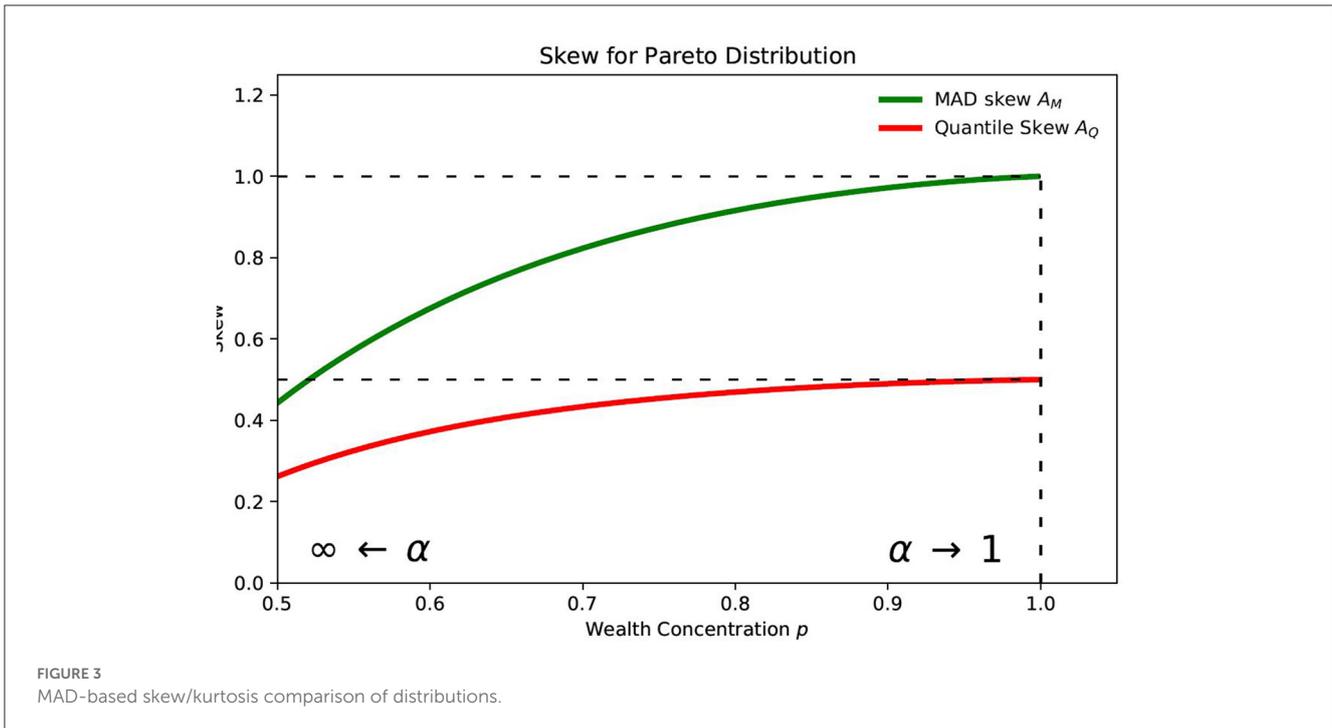
FIGURE 1
Skew for Pareto distribution.

FIGURE 2
Kurtosis for Pareto distribution.

This $p + q$ principle with $p > q$ corresponds to a Pareto distribution with a tail index α satisfying

$$\alpha = \frac{\log(p)}{\log(p) - \log(q)}$$

For example, the $60 - 40$ rule ($p = 0.6$, $q = 0.4$) has $\alpha = 2.260$ whereas $80 - 20$ rule ($p = 0.8$, $q = 0.2$) corresponds to $\alpha = \log(5)/\log(4) \approx= 1.161$. If we take even larger p as in $95 - 5$ rule ($p = 0.95$, $q = 0.05$ we get $\alpha = \log(0.05)/\log(0.95) \approx 1.017$. It is easy to prove that $\partial\alpha/\partial p < 0$ and, therefore, α decreases as

FIGURE 3
MAD-based skew/kurtosis comparison of distributions.

p increases. In particular, $\alpha \searrow 1$ as $p \mapsto 1$. Larger values of p correspond to higher concentrations of wealth.

Figure 1 demonstrates that the MAD-based skewness approaches its asymptotic value of $+1$ more rapidly than the quantile-based skewness approaches its asymptotic value of 0.5. This behavior demonstrates that the MAD-based skewness A_M is better able to explore the extreme distribution of wealth and income in the highest brackets than using the quantile-based skewness.

It is important to note that in the case of $\alpha < 2$, we cannot use classical measures for standard deviation, skewness, and kurtosis. By contrast, the MAD-based measures could be computed for these lower values of α and compared with the corresponding quantile-based measures.

As shown in Figure 2, the quantile-based kurtosis approaches its asymptotic value of 2.17 (from Equation A15 in Appendix) more rapidly than MAD-based kurtosis approaches its asymptotic value of 1. Despite this behavior in the kurtosis, we believe that skewness is more widely used and more revealing as a measure of the concentration of wealth and income at the highest brackets. The MAD-based kurtosis has the added practical advantage of having a maximum value of 1.

10. Comparison of distributions

Using MAD-based alternatives for deviation, skewness and kurtosis gives us additional ways to compare distributions. The MAD-based skewness A_M and MAD-based kurtosis T_M are always in the range $-1 \leq A_M \leq 1$ and $0 \leq T_M \leq 1$, respectively. This is in contrast to classical and quantile-based measures that can be unbounded. Therefore, with MAD-based measures, we have a way

to directly compare any two distributions in terms of their MAD-based skewness and kurtosis. In Figure 3, we plot both skewness and kurtosis for the distributions considered.

11. Conclusion

This paper considered MAD (about median)-based alternative metrics for classical standard deviation, skewness, and kurtosis. These MAD-based measures are shift-invariant. The MAD-based measures for skewness and kurtosis are also scale invariant. They can be computed from the corresponding left and right sub-ranges and require the existence of first-order moments only. The mathematical expressions for these measures are similar to those in quantile-based measures but involve computing means as opposed to quantiles. The resulting expressions can be interpreted as average values distance in sub-ranges from their respective medians.

In terms of practical applications, it is widely recognized that the median can be a better measure of centrality when the mean is overly influenced by outlier concentrations at the high end. The median captures the data centrality in many distributions with concentrations at the very high end, such as wealth and income. Building on the recognition that the median is the preferred metric in such applications, we go further in proposing MAD-based (about Median) metrics that give additional information and insight on the concentrations in the highest quantile.

For any distribution, the proposed MAD-based expressions for skewness and kurtosis are shown to be in the range $[-1, 1]$ and $[0, 1]$, respectively. The proposed MAD-based alternative measures provide a universal

scale to compare skewness and kurtosis across different data sets.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SK proposed the initial idea and focused on examples and discussion. EP focused on analytical work. Both authors have read and approved the final manuscript.

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2023.1206537/full#supplementary-material

References

1. Pham-Gia T, Hung TL. The mean and median absolute deviations. *J Math Comput Modell.* (2001) 34:921–36. doi: 10.1016/S0895-7177(01)00109-1

2. Gorard S. Revisiting a 90-year-old debate: the advantages of the mean deviation. *Brit J Educ Stud.* (2005) 53:417–30. doi: 10.1111/j.1467-8527.2005.00304.x

3. Farebrother RW. The historical development of the L_1 and L_∞ estimation methods. In: Dodge Y, editor. *Statistical data Analysis Based on the L_1-norm and Related Topics.* Amsterdam: North-Holland (1987). p. 37–63.

4. Portnoy S, Koenker R. The Gaussian hare and the Laplacian tortoise: computability of square-error versus abolute-error estimators. *Stat Sci.* (1997) 2:279–300.

5. Dodge Y. *Statistical Data Analysis Based on the L_1 Norm and Related Topics.* Amsterdam: North-Holland (1987).

6. Elsayed KMT. Mean absolute deviation: analysis and applications. *Int J Bus Stat Anal.* (2015) 2:63–74. doi: 10.12785/ijbsa/020201

7. Gorard S. An absolute deviation approach to assessing correlation. *Brit J Educ Soc Behav Sci.* (2015) 53:73–81. doi: 10.9734/BJESBS/2015/11381

8. Gorard S. Introducing the mean deviation "effect" size. *Int J Res Method Educ.* (2015) 38:105–14. doi: 10.1080/1743727X.2014.920810

9. Rousseeuw PJ, Croux C. Alternatives to the median absolute deviation. *J Am Stat Assoc.* (1993) 88:1273–83.

10. Bloomfield P, Steiger WL. *Least Absolute Deviations: Theory, Applications and Algorithms.* Boston, MA: Birkhauser (1983).

11. Shad SM. On the minimum property of the first absolute moment. *Am Stat.* (1969) 23:27.

12. Schwertman NC, Gilks AJ, Cameron J. A simple noncalculus proof that the median minimizes the sum of the absolute deviations. *Am Stat.* (1990) 44:38–41.

13. Upton G, Cook I. *Understanding Statistics.* Oxford: Oxford University Press (1996).

14. Zwillinger D, Kokoska S. *CRC Standard Probability and Statistics Tables and Formulae.* New York, NY: CRC Press (2000).

15. MacGillivray HI. Skewness and asymmetry. *Ann Stat.* (1986) 14:994–1011.

16. Groeneveld RA, Meeden G. Measuring skewness and kurtosis. *J R Stat Soc Ser D.* (1984) 33:391–99.

17. Habib E. Mean absolute deviation about median as a tool of exploratory data analysis. *Int J Res Rev Appl Sci.* (2012) 11:517–23.

18. Bowley AL. *Elements of Statistics.* 4th ed. New York, NY: Charles Scribner (1920).

19. Yule GU. *An Introduction to the Theory of Statistics.* London: C. Griffin, Limited (1912).

20. Johnson NL, Kotz S. *Distributions in Statistics.* New York, NY: J. Wiley (1970).

21. Moors JJ. A quantile alternative to kurtosis. *J R Stat Soc Ser D.* (1988) 37:25–32.

22. Feller JJ. *Probability Theory and Applications.* New York, NY: J. Wiley (1956).

23. Huber PJ. *Robust Statistics.* New York, NY: J. Wiley (2009).

24. Bennett N, Hays D, Sullivan B. *The Wealth of Households: 2019.* Washington, DC: Department of Commerce (2019). p. 1–6.