



OPEN ACCESS

EDITED BY

Charles K. Chui,
Stanford University, United States

REVIEWED BY

Zygmunt Warsza,
Łukasiewicz Research Network Industrial
Research Institute for Automation and
Measurements, Poland
Diganta Mukherjee,
Indian Statistical Institute, India

*CORRESPONDENCE

Eugene Pinsky
✉ epinsky@bu.edu

RECEIVED 08 August 2023

ACCEPTED 14 November 2023

PUBLISHED 07 December 2023

CITATION

Wang Y and Pinsky E (2023) Geometry of
deviation measures for triangular distributions.
Front. Appl. Math. Stat. 9:1274787.
doi: 10.3389/fams.2023.1274787

COPYRIGHT

© 2023 Wang and Pinsky. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Geometry of deviation measures for triangular distributions

Yuhe Wang and Eugene Pinsky*

Computer Science Department, MET College, Boston University, Boston, MA, United States

Triangular distributions are widely used in many applications with limited sample data, business simulations, and project management. As with other distributions, a standard way to measure deviations is to compute the standard deviation. However, the standard deviation is sensitive to outliers. In this paper, we consider and compare other deviation metrics, namely the mean absolute deviation from the mean, the median, and the quantile-based deviation. We show the simple geometric interpretations for these deviation measures and how to construct them using a compass and a straightedge. The explicit formula of mean absolute deviation from the median for triangular distribution is derived in this paper for the first time. It has a simple geometric interpretation. It is the least volatile and is always better than the standard or mean absolute deviation from the mean. Although greater than the quantile deviation, it is easier to compute with limited sample data. We present a new procedure to estimate the parameters of this distribution in terms of this deviation. This procedure is computationally simple and may be superior to other methods when dealing with limited sample data, as is often the case with triangle distributions.

KEYWORDS

triangular distribution, mean absolute deviation, quartile deviations, probability distributions, geometry of deviations

1 Introduction

Triangular distributions $Tr(a, b, c)$ are a three-parameter distribution with a minimum value a , a maximum value b , and a most likely (mode) value c [1]. Unlike the uniform distribution, triangle distributions include the most likely value c in addition to the minimum and maximum values and allow for incorporating skewness and asymmetry. These distributions are widely used in business simulations, risk and project management, Monte-Carlo simulations, and other areas, especially when dealing with limited sample data [2–4] (for a historical review, see [1]).

As noted in [5], several authors [1, 6] have suggested using this distribution instead of the beta distribution since for beta distribution, there are difficulties in the maximum likelihood estimation of parameters and their interpretation. Johnson [6] and Johnson and Kotz [7] dealt with neglected applications of this distribution as an alternative to the beta distribution, which suffers from difficulties in its maximum likelihood parameter estimation. Triangular distributions are recommended when the underlying distribution is unknown, but a minimal value, some maximal value, and a most likely value are available [1].

This paper focuses on computing and comparing the deviation measures for such distributions, both geometrically and algebraically. Historically, the most widely used deviation measure has been the standard deviation [8]. In computing this metric, we square the distances from the mean. This usage of the L_2 norm is convenient in differentiation, optimization, and estimation [9]. However, this metric has some disadvantages, such as sensitivity to outliers. An alternative is to use the mean absolute deviation (the L_1 -norm)

and measure deviations (MAD) from a central point such as mean μ [denoted by $H(\mu)$] or median M [denoted in this paper by $H(M)$]. On the other hand, using L_1 metric allows one to introduce performance metrics that do not require the computation of higher-order moments (e.g. [10]).

One of the contributions of this paper is an explicit formula for $H(M)$, the mean absolute deviation about the median. The formula has a simple, intuitive, and geometric interpretation. To our knowledge, this is the first paper deriving an explicit formula for $H(M)$ for the triangular distribution. For this distribution, deviation measures such as standard deviation, mean absolute deviation (about median), and quartile deviation have simple geometric interpretations and can all be constructed by a ruler and compass. The mean absolute deviation about the median $H(M)$ is the least volatile and is always lower than the standard deviation or the MAD deviation about the mean. Although it is higher than the quartile deviation, it is more practical. We often have very limited data for triangular distributions, and accurate quartile estimates are problematic. Using $H(M)$, we present a new non-iterative procedure to estimate the parameters of this distribution. This procedure is very simple computationally and is more robust than direct estimation based on the sample mean.

This paper is organized as follows. In Section 2, we introduce triangular distributions, compare mean and median, and present geometric interpretations for mean, median, and standard deviation. In Section 3 we derive a general expression for computing mean absolute deviations for triangular distributions. From this expression, we compute the mean absolute deviations $H(\mu)$ and $H(M)$ about mean and median and provide a simple geometric interpretation for these measures. In Section 4, we introduce a new estimation procedure based on mean absolute deviation $H(M)$ and illustrate it with a numerical example. In Section 5, we compare the exact values of tail probabilities with some estimates using deviations. In Section 6, we derive the formula for the quartile deviation. Finally, in Section 7 we present a detailed comparison of deviation measures.

2 Triangular distributions

We start with definitions and notations. Suppose X is distributed according to triangular distribution on the range $[a, b]$ and mode parameter $c \in [a, b]$. We will write this as $X \sim \text{Tr}(a, b, c)$. The density function $f(x)$ and the cumulative distribution function $F(x)$ is given by

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{for } a \leq x \leq c, \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{for } c < x \leq b \end{cases} \quad F(x) = \begin{cases} \frac{(x-a)^2}{(b-a)(c-a)} & \text{for } a \leq x \leq c, \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{for } c < x \leq b \end{cases} \quad (1)$$

The mean μ and standard deviation σ are:

$$\mu = (a + b + c)/3 \text{ and } \sigma = \sqrt{a^2 + b^2 + c^2 - ab - ac - bc}/3\sqrt{2} \quad (2)$$

Let $F^{-1}(\cdot)$ be the quantile function defined by $F^{-1}(t) = \inf\{x : F(x) \geq t\}$ with $t \in [0, 1]$. From the definition of the

cumulative distribution function, we find

$$x = F^{-1}(t) = \begin{cases} a + \sqrt{t(b-a)(c-a)} & a \leq x \leq c \\ b - \sqrt{(1-t)(b-a)(b-c)} & c < x \leq b \end{cases} \quad (3)$$

Therefore, for the median $M = F^{-1}(1/2)$ we have

$$\begin{cases} (b - M) = \frac{\sqrt{2}}{2} \sqrt{(b-a)(b-c)} & c \leq (a + b)/2 \\ (M - a) = \frac{\sqrt{2}}{2} \sqrt{(b-a)(c-a)} & c \geq (a + b)/2 \end{cases} \quad (4)$$

We will find it convenient to consider a standard (“normalized”) variable under min-max transformation $X \mapsto (X - a)/(b - a)$. Then, X has a triangular distribution with $a = 0, b = 1$ and $a \leq c \leq b$. For this standard X , the mean $\mu = (1 + c)/3$, standard deviation $\sqrt{1 + c^2 - c}/3\sqrt{2}$ whereas the median $M = 1 - \sqrt{2(1 - c)}/2$ for $c \leq 1/2$ and $M = \sqrt{c/2}$ for $1/2 \leq c \leq 1$.

In Table 1, we compute mean, median and standard deviation for standard triangle distribution ($a = 0, b = 1$) for $c = 0, c = 1/4, c = 1/2, c = 3/4$, and $c = 1$.

The case $c = 0$ corresponds to the distribution $X = |U_1 - U_2|$ of the absolute difference between two independent random variables U_1 and U_2 with standard uniform distribution. By contrast, the symmetric case $c = 1/2$ corresponds to the distribution $X = (U_1 + U_2)/2$ of the mean of the sum of two standard uniform variables U_1 and U_2 .

The mean μ , the median M , the standard deviation σ , and other deviation measures discussed in this paper have simple geometric interpretations and can be constructed using a straightedge and compass. For their constructions, we will assume that we know how to do the following:

1. construct a segment of unit length
2. copy segments and angles
3. construct $(l_1 + l_2)$ and $(l_1 - l_2)$ from segments of lengths l_1 and l_2
4. divide length l into 2 (bisection) equal or 3 equal parts (trisection)
5. construct $30^\circ/60^\circ$ and $45^\circ/45^\circ$ right triangles with a side (e.g., hypotenuse) of length l . This allows computing $l\sqrt{2}, l/\sqrt{2}, l\sqrt{3}$, and $l/\sqrt{3}$
6. construct a right triangle with sides l_1 and l_2
7. compute the geometric mean $\sqrt{l_1 l_2}$ of two segments with lengths l_1 and l_2 . This construction is illustrated in Figure 1.

We can now discuss the geometric construction and interpretation of some measures for this distribution.

2.1 Geometric construction and interpretation of the mean

To construct μ , it will be sufficient to construct $(\mu - a)$. From Equation (2) we have $(\mu - a) = (c - a)/3 + (b - a)/3$. This gives us the following construction for $(\mu - a)$:

1. trisect $(b - a)$ to compute $l_1 = (b - a)/3$

TABLE 1 Mean, median, and standard deviation.

| c | 0 | 0.25 | 0.5 | 0.75 | 1 |
|----------|-------------------------------------|-------------------------------------|------------------------------------|-------------------------------------|-----------------------------------|
| μ | $\frac{1}{3} \approx 0.33$ | $\frac{5}{12} \approx 0.42$ | 0.5 | $\frac{7}{12} \approx 0.58$ | $\frac{2}{3} \approx 0.66$ |
| M | $\frac{2-\sqrt{2}}{2} \approx 0.29$ | $\frac{4-\sqrt{6}}{4} \approx 0.39$ | 0.5 | $\frac{\sqrt{6}}{4} \approx 0.61$ | $\frac{\sqrt{2}}{2} \approx 0.71$ |
| σ | $\frac{\sqrt{2}}{6} \approx 0.24$ | $\frac{\sqrt{26}}{24} \approx 0.21$ | $\frac{\sqrt{6}}{12} \approx 0.20$ | $\frac{\sqrt{26}}{24} \approx 0.21$ | $\frac{\sqrt{2}}{6} \approx 0.24$ |

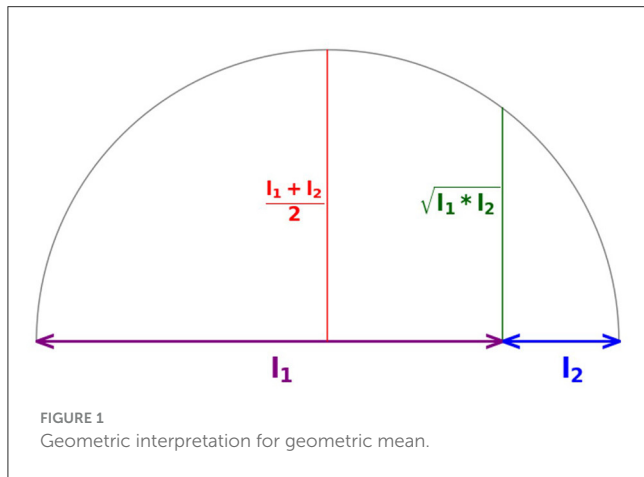


FIGURE 1 Geometric interpretation for geometric mean.

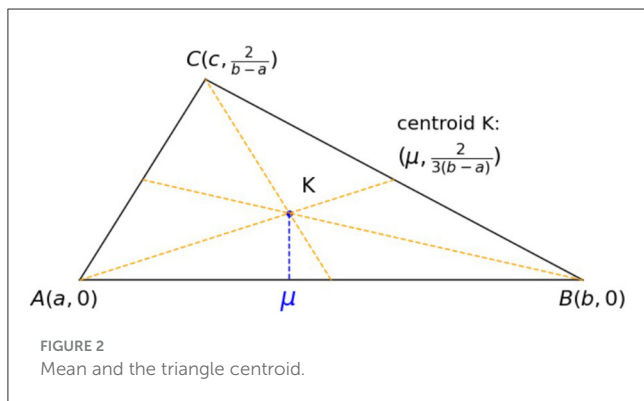


FIGURE 2 Mean and the triangle centroid.

2. trisect $(c - a)$ to compute $l_2 = (c - a)/3$
3. compute $(\mu - a) = l_1 + l_2$

The mean μ has the following interesting interpretation: Consider a triangle with vertices (x_1, y_1) , (x_2, y_2) , and (x_3, y_3) . Consider the centroid K of such a triangle: the point at which the three medians of the triangle intersect as shown in Figure 2. This centroid is the point with coordinates $(x_1 + x_2 + x_3)/3, (y_1 + y_2 + y_3)/3$. To construct this point, we bisect the sides of the triangle and draw lines from vertices to the midpoints of the opposite sides. These three lines meet at the single point K . By the centroid theorem, the point K is $2/3$ of the distance from the vertex to the midpoint of the sides. If we use the geometric interpretation of the density as a triangle with vertices $(a, 0)$, $(b, 0)$, and $(c, 2/(b - a))$, then the x -coordinate of its centroid is the mean μ .

Next, we establish the relationship between the mean μ and the midpoint $(a + b)/2$. If $c \leq (a + b)/2$, then $3c \leq (a + b + c)$ or $c \leq \mu$. On the other hand, $\mu = (a + b + c)/3 \leq ((a + b) + (a + b)/2)/3 =$

$(a + b)/2$ and thus, $c \leq \mu \leq (a + b)/2$. Therefore,

$$0 \leq c \leq (a + b)/2 \implies \mu \geq c \quad \text{and} \quad (a + b)/2 \leq c \leq b \implies \mu \leq c \tag{5}$$

2.2 Geometric construction and interpretation of the median

The median M has a different geometric interpretation. We consider the case $c \leq (a + b)/2$. To construct M , it will be sufficient to construct $(M - a)$. We proceed as follows:

1. compute the geometric mean $l_1 = \sqrt{(b - a)(b - c)}$
2. (any other) side of this triangle has the length $(M - a) = l_1 \sqrt{2}/2$

This is illustrated in Figure 3.

For the case $c \geq (a + b)/2$, we construct $(b - M)$ by similar steps:

1. compute the geometric mean $l_1 = \sqrt{(b - a)(c - a)}$
2. (any other) side of this triangle has a length $(b - M) = l_1 \sqrt{2}/2$

Next, we establish the relationship between the median M and the midpoint $(a + b)/2$. There are two cases to consider:

- Case 1: $c \leq (a + b)/2$. In this case $(b - a) \leq 2(b - c)$ and from Equation (4) and we have

$$M = b - \sqrt{(b - a)(b - c)}/\sqrt{2} \geq b - (b - c) = c$$

- Case 2: $c \geq (a + b)/2$. In this case, $(b - a) \leq 2(c - a)$ and we have

$$M = a + \sqrt{(b - a)(c - a)}/\sqrt{2} \leq a + (c - a) = c$$

Therefore, we obtain the following relationship:

$$a \leq c \leq (a + b)/2 \implies M \geq c \quad \text{and} \quad (a + b)/2 \leq c \leq b \implies M \leq c \tag{6}$$

The case $c \leq (a + b)/2$ is illustrated in Figure 4.

2.3 Relationship of mean and median

We show that for $a \leq c \leq (a + b)/2$ we have $c \leq M \leq \mu$ and for $(a + b)/2 \leq c \leq b$ we have $\mu \leq M \leq c$. There are two cases to consider:

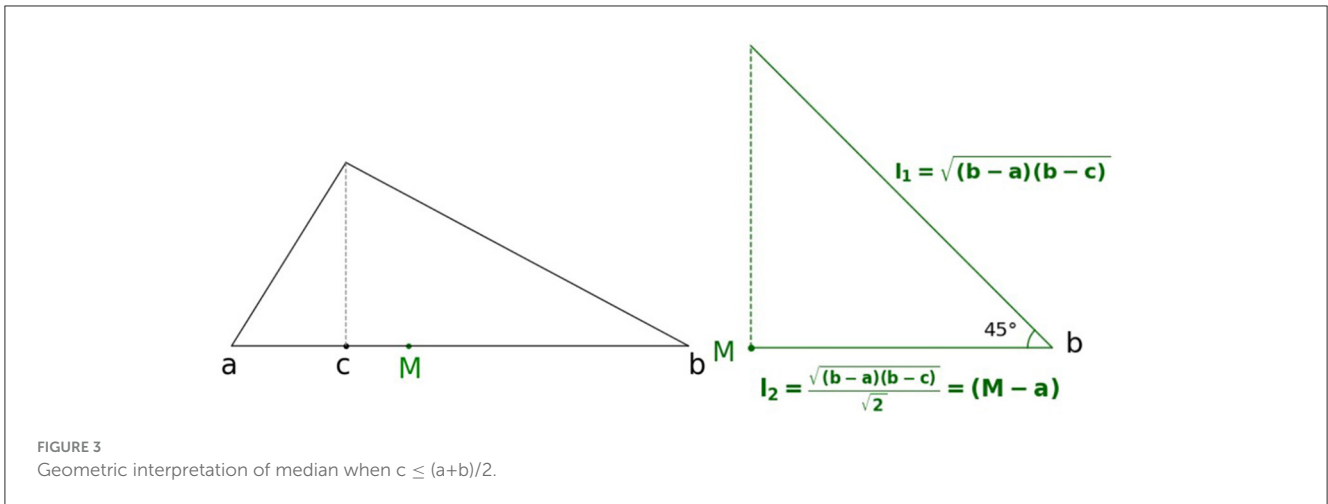


FIGURE 3 Geometric interpretation of median when $c \leq (a+b)/2$.

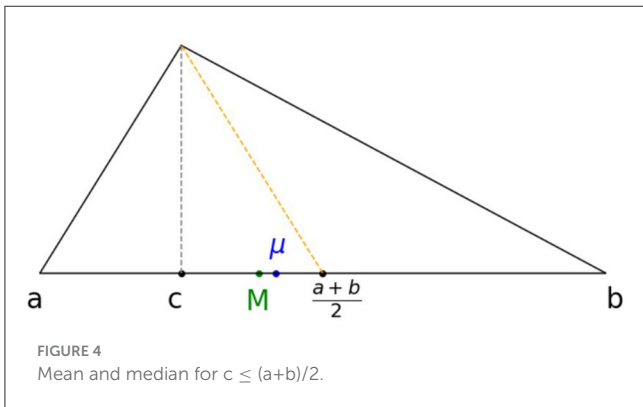


FIGURE 4 Mean and median for $c \leq (a+b)/2$.

- Case 1: $a \leq c \leq (a + b)/2$. In this case, $(b - a)(b - c) \geq (b - a)^2/2$ and from the Equation (4)

$$M = \frac{b^2 - \frac{1}{2}(b-a)(b-c)}{b + \frac{\sqrt{2}}{2}\sqrt{(b-a)(b-c)}} \leq \frac{1}{2} \cdot \frac{(b^2 + ab + bc - ac)}{b + \frac{1}{2}(b-a)} \leq \frac{(a+b+c)}{3} = \mu \tag{7}$$

- Case 2: $(a + b)/2 \leq c \leq b$. If we consider the distribution $Y = (a + b) - X$ then by symmetry from the above Equation (7) we have $\mu \leq M \leq c$.

Combining these results with Equation (6) we obtain the following relationship between the mean μ , the median M and the mode c :

$$\begin{aligned} a \leq c \leq (a + b)/2 &\implies a \leq M \leq \mu \leq c \\ (a + b)/2 \leq c \leq b &\implies c \leq \mu \leq M \leq b \end{aligned} \tag{8}$$

2.4 Geometric construction and interpretation of the standard deviation

To derive a geometric interpretation for σ , we can re-write the expression for σ in Equation (2) as follows

$$\sigma = \frac{1}{3} \cdot \frac{1}{\sqrt{2}} \sqrt{(b - a)^2 - (c - a)(b - c)} \tag{9}$$

From this equation, we immediately obtain $\sigma \leq \sqrt{2}(b - a)/6$. For the case $c = a$ or $c = b$, we have the equality $\sigma = \sqrt{2}(b - a)/6$.

The above Equation (9) suggests the following geometric construction and interpretation for the standard deviation σ are shown in Figure 5.

1. compute the geometric mean $l_2 = \sqrt{(c - a)(b - c)}$
2. construct the right triangle T with hypotenuse $(b - a)$ as side l_1 .
3. compute the length $l_3 = \sqrt{(b - a)^2 - l_1^2}$ of the other side of T
4. compute the length $l_4 = l_3/\sqrt{2}$
5. trisect l_3 to compute $\sigma = l_4/3$

3 Mean absolute deviation

We start with preliminary definitions. Consider a real-valued random variable X with density $f(x)$, finite mean and cumulative distribution function $F(x)$. We use M , $\mu = E(X)$ and σ to denote median, mean and standard deviation of X respectively. For any p , we define the mean absolute deviation of X from p as

$$H(p) = E(|X - p|) = \int_X |x - p|f(x) dx \tag{10}$$

It can be interpreted as the average distance of values of X to the point p . If $p = \mu$, then $H(\mu)$ is the mean absolute deviation from the mean μ . If we take $p = M$, then $H(M)$ is the mean absolute deviation from the median. Both are denoted as MAD (mean absolute deviation) in the statistical literature, leading to some confusion [9].

It can be shown (details are presented in the Appendix) that for any distribution $\sigma \geq H(\mu) \geq H(M)$. Of the three metrics

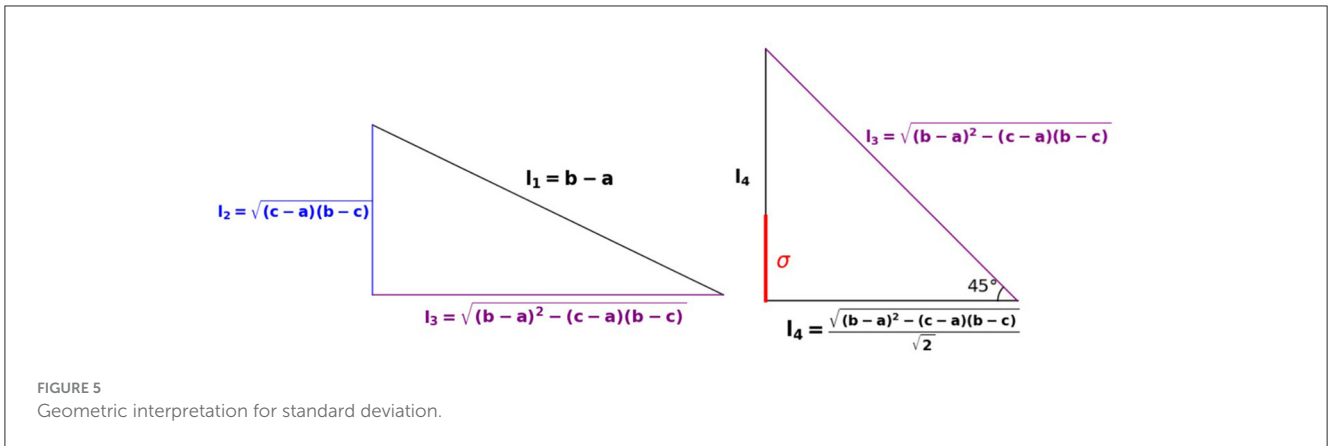


FIGURE 5 Geometric interpretation for standard deviation.

to measure deviations, namely $H(M)$, $H(\mu)$, and σ , the MAD metric $H(M)$ has the lowest value. In the computation of σ , we square distances (from the mean), and this inflates the impact of outliers. By contrast, we use these distances directly to compute mean absolute deviations. In addition, it is often easier to interpret mean absolute deviations than the standard deviation.

We illustrate this with a simple example. 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, $\mu = M = (a + b)/2$, and $\sigma = \sqrt{3}(b - a)/6$. Since this distribution is symmetric, the MAD deviations $H(M) = H(\mu)$ and are easily computed $H(M) = (b - a)/4$ (e.g. [10]). The MAD value $H(M) = (b - a)/4$ is easy to interpret: it represents the average distance of X from its median $M = (a + b)/2$. However, it is more difficult to find an easy interpretation for standard deviation $\sigma = \sqrt{3}(b - a)/6$.

In the Appendix, we provide computational details on computing $H(\mu)$ and $H(M)$. These are summarized below:

For triangular distribution, the mean absolute deviation from the mean $H(\mu)$ is

$$H(\mu) = \begin{cases} \frac{2(a+c-2b)^3}{81(a-b)(b-c)} & a \leq c \leq (a+b)/2 \\ \frac{2(b+c-2a)^3}{81(a-b)(a-c)} & (a+b)/2 \leq c \leq b \end{cases} \quad (11)$$

For the standard triangular distribution with $a = 0, b = 1$ this simplifies to:

$$H(\mu) = \begin{cases} \frac{2}{3(1-c)} \left(\frac{2-c}{3}\right)^3 & 0 \leq c \leq 1/2 \\ \frac{2}{3c} \left(\frac{1+c}{3}\right)^3 & 1/2 \leq c \leq 1 \end{cases} \quad (12)$$

For the mean absolute deviation $H(M)$ from the median we obtain

$$H(M) = \begin{cases} \frac{(b-a)+(b-c)-\sqrt{2(b-a)(b-c)}}{3} & a \leq c \leq (a+b)/2 \\ \frac{(b-a)+(c-a)-\sqrt{2(b-a)(c-a)}}{3} & (a+b)/2 \leq c \leq b \end{cases} \quad (13)$$

For the standard triangular distribution with $a = 0, b = 1$, this simplifies to

$$H(M) = \begin{cases} \frac{1}{3}(2 - c - \sqrt{2(1-c)}) & 0 \leq c \leq 1/2 \\ \frac{1}{3}(1 + c - \sqrt{2c}) & 1/2 \leq c \leq 1 \end{cases} \quad (14)$$

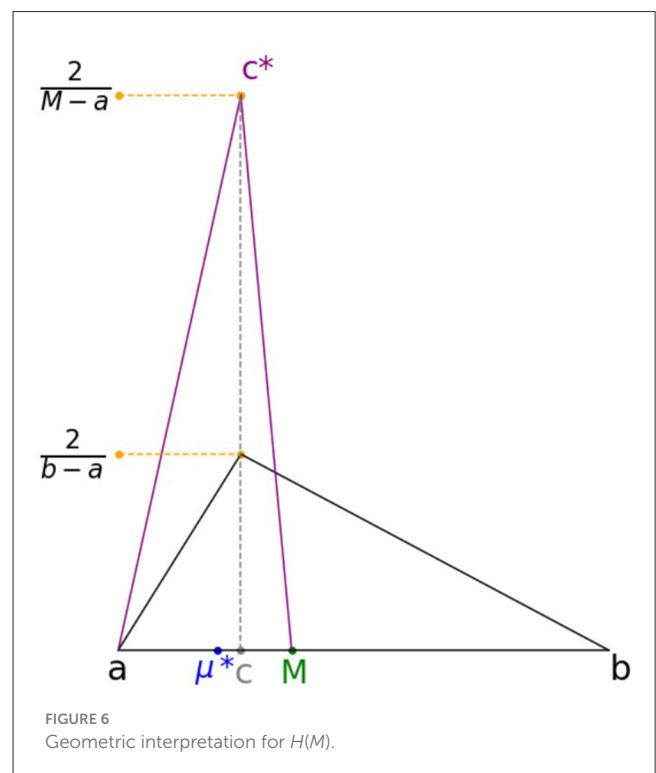


FIGURE 6 Geometric interpretation for $H(M)$.

3.1 Geometric construction and interpretation of MAD

The Equation (13) gives us the following procedure for the geometric construction of $H(M)$. We consider two cases:

- Case 1: $a \leq c \leq (a + b)/2$. We construct $H(M)$ as follows:
 1. construct the geometric mean $l_1 = \sqrt{(b - a)(b - c)}$
 2. compute length $l_2 = l_1\sqrt{2}$.
 3. compute $l_3 = (b - a) + (b - c) + l_2$
 4. trisect l_3 to compute $H(M) = l_3/3$
- Case 2: $c \geq (a + b)/2$. We construct $H(M)$ by similar steps:
 1. construct the geometric mean $l_1 = \sqrt{(b - a)(c - a)}$

2. compute its hypotenuse $l_2 = l_1\sqrt{2}$.
3. compute $l_3 = (b - a) + (c - a) + l_2$
4. trisect l_3 to compute $H(M) = l_3/3$

The mean absolute deviation $H(M)$ can be interpreted as follows. For $c \leq (a + b)/2$ from Case 1 above, we can write $H(M) = M - (a + c + M)/3$. The term $(a + M + c)/3$ is the mean of the triangular distribution $Y \sim \text{Tr}(a, M, c)$ with parameters a, M , and c . Therefore, $H(M)$ can be interpreted as the distance between the median M and the mean of the triangular distribution $\text{Tr}(a, M, c)$.

Similarly, for $c \geq (a + b)/2$ from Case 2 above, we can write $H(M) = (b + c + M)/3 - M$. The term $(a + M + c)/3$ is the mean of the triangular distribution $Z \sim \text{Tr}(M, c, b)$ with parameters M, c , and b . Therefore, $H(M)$ can be interpreted as the distance between the median M and the mean of the triangular distribution $\text{Tr}(M, c, b)$. This is illustrated in the Figure 6.

In Table 2, we present additional results for comparing standard and mean absolute deviations for normalized triangular distributions for $c = \{0, 0.25, 0.5, 0.75, 1\}$. These comparisons are further illustrated in Figure 7.

As a specific example, consider the symmetric triangular distribution with $a = 0, b = 1$ and $c = 1/2$. For this distribution, $\mu = M = 1/2$. The standard deviation $\sigma = \sqrt{2}/6$. The mean absolute deviations $H(\mu) = H(M) = 1/6$. The case $c = 1/2$ corresponds to the distribution $X = (U_1 + U_2)/2$ of the mean of the sum of two independent standard uniform variables U_1 and U_2 . For these uniform variables, the mean is $\mu(U_1) = \mu(U_2) = 1/2$, standard deviation $\sigma(U_1) = \sigma(U_2) = 1/(2\sqrt{3})$ and MAD deviation $H(U_1, 1/2) = H(U_2, 1/2) = 1/4$. We note that $\sigma^2(X) = \sigma^2(U_1/2) + \sigma^2(U_2/2)$ but $H(1/2) \neq H(U_2/2, 1/2) + H(U_2/2)$. From this, we see that, unlike variances, the mean absolute deviation is not additive for the sum of independent random variables. We also note that for the triangular distribution, its standard deviation $\sigma(\text{triangle}) = \sqrt{2}/6$ is much larger than the standard deviation of a (normalized) uniform $\sigma(\text{uniform}) = \sqrt{3}/6$.

4 Estimation of parameters with mean absolute deviation

In this section, we address the issue of estimating the parameters of this distribution. We could consider several methods to estimate parameters a, b , and c such as maximum likelihood estimation [5, 11], quantiles [4, 12], method of moments or other widely used methods in statistics.

However, we must remember that we assume limited sample data for these distributions. Therefore, the computations based on quantiles may not be very accurate or practical. There are difficulties in using maximum likelihood estimation for triangular distributions as well [13, 14]. One difficulty is that the original density function $f(x)$ in Equation (1) has a “sharp” corner at point c where it does not have a derivative.

4.1 Mean-based and midrange-based estimation of parameters

The simple relationship $\mu = (a + b + c)/3$ in Equation (2) for the mean μ suggests the following simple procedure to estimate the parameters. If we have a set of n data points x_1, \dots, x_n then we have the following:

1. compute $a^* = \min\{x_i\}$ and $b^* = \max\{x_i\}$
2. compute the sample mean $\mu^* = (x_1 + \dots + x_n)/n$
3. compute the mode $c^* = 3\mu^* - (a^* + b^*)$

An even simpler procedure would be to use the mid-range estimator for μ (i.e., assume a symmetric triangle distribution), and proceed as follows:

1. compute $a^* = \min\{x_i\}$ and $b^* = \max\{x_i\}$
2. estimate the sample mean by mid-range $\mu^* = (a^* + b^*)/2$
3. compute the mode $c^* = 3\mu^* - (a^* + b^*)$

However, the mean μ^* is sensitive to outlier values. However, the mean is more sensitive to outlier values than the median. We, therefore, suggest an alternative method based on the sample median M^* and $H(M^*)$.

4.2 MAD (around median)-based estimation of parameters

Recall that for the median M , we have shown $c \leq (a + b)/2 \iff M \leq c$. Combining this with expressions for $H(M)$ in Equation (13), we obtain

$$H(M) = \begin{cases} (2M - a - c)/3 & M \leq (a + b)/2 \\ (b + c - 2M)/3 & M \geq (a + b)/2 \end{cases} \quad (15)$$

The above results suggest the following estimation procedure:

1. compute $a^* = \min\{x_i\}$ and $b^* = \max\{x_i\}$
2. compute the sample median M^*
3. compute the mean absolute deviation $H(M^*)$ from the (estimated sample) median M^*

$$H(M^*) = \frac{1}{n} \sum_{i=1}^n |x_i - M^*| \quad (16)$$

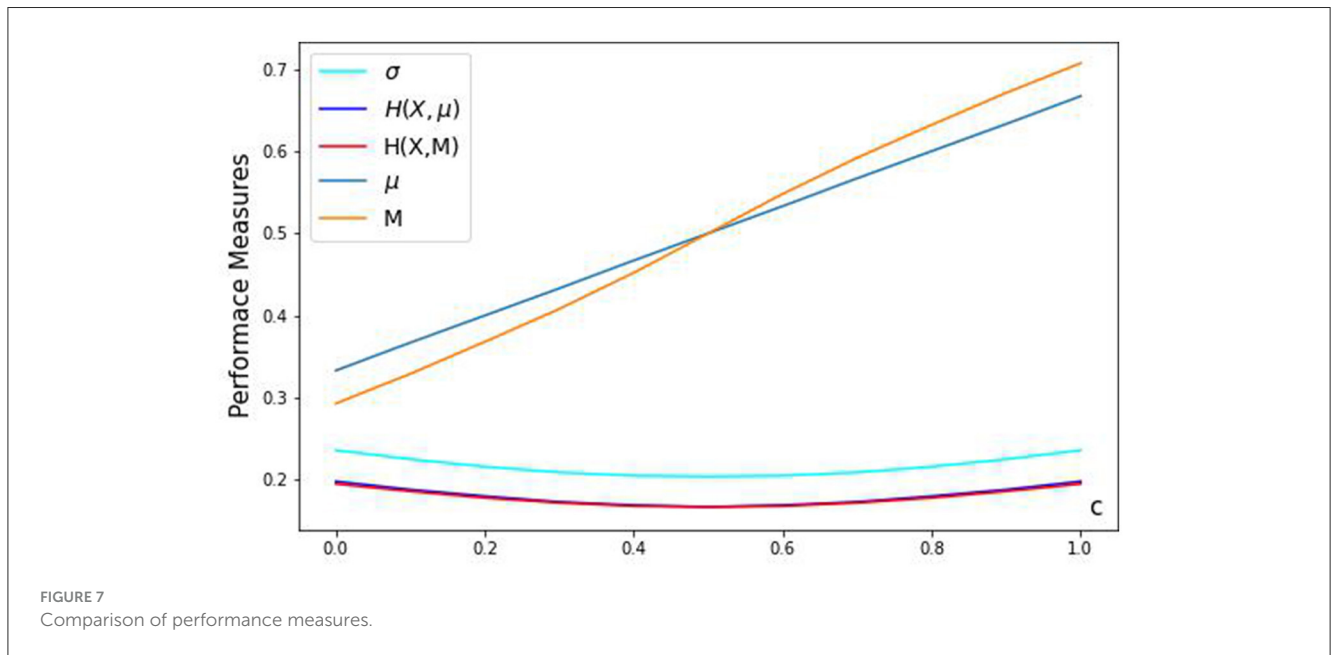
4. from Equation (15) compute the mode c^* as follows

$$c^* = \begin{cases} 2M^* - 3H(M^*) - a^* & M^* \leq (a^* + b^*)/2 \\ 2M^* + 3H(M^*) - b^* & M^* \geq (a^* + b^*)/2 \end{cases} \quad (17)$$

This procedure is more stable and less sensitive to outliers than the simple procedure based on the means. This is illustrated by a numerical example below.

TABLE 2 A comparison of standard and mean absolute deviations.

| c | 0 | 0.25 | 0.5 | 0.75 | 1 |
|----------|-------------------------------------|--------------------------------------------|------------------------------------|--------------------------------------------|-------------------------------------|
| μ | $\frac{1}{3} \approx 0.33$ | $\frac{5}{12} \approx 0.42$ | 0.5 | $\frac{7}{12} \approx 0.58$ | $\frac{2}{3} \approx 0.66$ |
| M | $\frac{2-\sqrt{2}}{2} \approx 0.29$ | $\frac{4-\sqrt{6}}{4} \approx 0.39$ | 0.5 | $\frac{\sqrt{6}}{4} \approx 0.61$ | $\frac{\sqrt{2}}{2} \approx 0.71$ |
| σ | $\frac{\sqrt{2}}{6} \approx 0.24$ | $\frac{\sqrt{26}}{24} \approx 0.21$ | $\frac{\sqrt{6}}{12} \approx 0.20$ | $\frac{\sqrt{26}}{24} \approx 0.21$ | $\frac{\sqrt{2}}{6} \approx 0.24$ |
| $H(\mu)$ | $(\frac{2}{3})^4 \approx 0.20$ | $\frac{1}{9} (\frac{7}{6})^3 \approx 0.18$ | $\frac{1}{6} \approx 0.17$ | $\frac{1}{9} (\frac{7}{6})^3 \approx 0.18$ | $(\frac{2}{3})^4 \approx 0.20$ |
| $H(M)$ | $\frac{2-\sqrt{2}}{3} \approx 0.20$ | $\frac{5-2\sqrt{2}}{12} \approx 0.18$ | $\frac{1}{6} \approx 0.17$ | $\frac{7-2\sqrt{6}}{12} \approx 0.18$ | $\frac{2-\sqrt{2}}{3} \approx 0.20$ |



4.3 A numerical example for parameter estimation

We present a numerical example of civil engineering data of 85 hauling times presented in [11]. The data is as follows:

- $X = \{4.79, 4.75, 5.40, 4.70, 6.50, 5.30, 6.00, 5.90, 4.80, 6.70, 6.00, 4.95, 7.90, 5.40, 3.50, 4.54, 6.90, 5.80, 5.40, 5.70, 8.00, 5.40, 5.60, 7.50, 7.00, 4.60, 3.20, 3.90, 5.90, 3.40, 5.20, 5.90, 4.40, 5.20, 7.40, 5.70, 6.00, 3.60, 6.20, 5.70, 5.80, 5.90, 6.00, 5.15, 6.00, 4.82, 5.90, 6.00, 7.30, 7.10, 4.73, 5.90, 3.60, 6.30, 7.00, 5.10, 6.00, 6.60, 4.40, 6.80, 5.60, 5.90, 5.90, 8.60, 6.00, 5.80, 5.40, 6.50, 4.80, 6.40, 4.15, 4.90, 6.50, 8.20, 7.00, 8.50, 5.90, 4.40, 5.80, 4.30, 5.10, 5.90, 4.70, 3.50, 6.80\}$

The mean of this data is $\mu^* = 5.69$, and the standard deviation is $\sigma^* = 1.16$. The Gaussian estimation is then $c^* = 5.69$.

The maximum likelihood estimation of parameters presented in Dorp and Kotz [5] gives the following results:

$$a^* = 2.87, \quad b^* = 8.80, \quad c^* = 5.80 \quad (18)$$

Let us estimate the parameters using the mean-based and MAD-based algorithm.

We start with the mean-based algorithm.

1. compute $a^* = \min\{x_i\} = 3.2$ and $b^* = \max\{x_i\} = 8.60$

2. compute the sample mean $\mu^* = 5.69$
3. compute the mode $c^* = 3\mu^* - (a^* + b^*) = 5.27$

Therefore, the mean-based algorithm gives the following:

$$a^* = 3.2, \quad b^* = 8.6, \quad c^* = 5.27 \quad (19)$$

If we consider using the mid-range estimation, then we proceed as follows:

1. compute $a^* = \min\{x_i\} = 3.2$ and $b^* = \max\{x_i\} = 8.60$
2. estimate the mean by mid-range $\mu^* = (a + b)/2 = 5.9$
3. compute the mode $c^* = 3\mu^* - (a^* + b^*) = 5.9$

Therefore, the mid-range estimation gives the following:

$$a^* = 3.2, \quad b^* = 8.6, \quad c^* = 5.9 \quad (20)$$

Next, we consider the MAD-based approach

1. compute $a^* = \min\{x_i\} = 3.2$ and $b^* = \max\{x_i\} = 8.60$
2. compute the sample median $M^* = 5.80$
3. compute the mean absolute deviation $H(M^*) = 0.88$ (from the median)
4. the midpoint $(a^* + b^*)/2 = 5.9$ is greater than M^* . Therefore, from Equation (17) we compute the mode $c^* = (2M^* - 3H(M^*)) - a^* = 5.75$

TABLE 3 Comparison of estimation for the numerical example.

| Method | a^* | b^* | c^* | σ^* |
|---------------------------------|-------|-------|-------|------------|
| Gauss | ... | ... | 5.69 | 1.16 |
| Maximum likelihood (MLE) | 2.87 | 8.80 | 5.80 | 1.21 |
| Mean-based | 3.20 | 8.60 | 5.27 | 1.11 |
| Mid-range based (symmetric PDF) | 3.20 | 8.60 | 5.90 | 1.10 |
| MAD-based | 3.20 | 8.60 | 5.75 | 1.10 |

Therefore, the MAD-based algorithm gives the following:

$$a^* = 3.2, \quad b^* = 8.6, \quad c^* = 5.75 \quad (21)$$

We summarize our results from Equations (18–21) in Table 3 where we also added the estimated standard deviation σ^* [computed from Equation (2)]

Examining Table 3, we see that both mean-based, symmetric PDF and MAD-based procedures give the same estimates for a^* and b^* . These are within about 10% relative errors when compared to the values computed by maximum likelihood in Equation (18). However, the estimate for mode c^* using mean-absolute deviation is much closer to the estimate obtained by maximum likelihood estimation (5.75 vs. 5.80 or about 1% relative error) than the other methods. The estimated standard deviation by MAD-based method is much lower than the one obtained by the maximum likelihood estimation (1.10 vs. 1.21). At the same time, the MAD-based procedure is computationally much simpler than fitting parameters using the maximum likelihood estimation.

We may consider other methods, based on quantiles. One such method presented in [12] assumes that the most likely value is given and solves for parameters a^* and b^* by numerically solving two quadratic equations corresponding to 5 and 95% quantiles. We could consider another approach: if we have the empirical quartiles Q_1^*, Q_3^* , we can compute the interquartile range $I = (Q_3^* - Q_1^*)$. One can then use the Friedman-Diaconis rule [15] to compute the bin width as $2I/n^{3/2}$, construct the empirical histogram and estimate c^* by the midpoint of the most frequent bin. From this, we can solve for a^* and b^* using Q_1^*, Q_3^* , and c^* .

However, these methods are pretty complex and require good estimates of the quantiles. In practice, the triangle distributions are often used when there is limited sample data available. For these reasons, the triangle distribution has been called a “lack of knowledge” distribution. Therefore, with small sample sizes, it is not feasible to consider many such methods. By contrast, the suggested MAD-based method for parameter estimation is much simpler computationally.

Let us now illustrate outlier detection using the above example. The simple classical approach is to use mean μ^* and standard deviation σ^* and look for outliers as values below the low bound $L = (\mu^* - 2\sigma^*)$ and above the upper bound $U = (\mu^* + 2\sigma^*)$. The MAD approach would be to use the estimated median M^* and MAD (around median) H^* and look for outlier values outside the low bound $L = (M^* - 2H^*)$ and above the upper bound $U = (M^* + 2H^*)$. We can also consider the simple non-parametric method

of Tukey using quartiles [16]. In Tukey’s method we compute the quartiles Q_1 and Q_3 , compute the interquartile range $I = (Q_3 - Q_1)$ and look for outlier values outside the low bound $L = Q_1 - 1.5I$ and above the upper bound $U = Q_3 + 1.5I$. We present a comparison of outlier detection for the above numerical example in Table 4.

We note that the lower and upper bounds for the above outlier detection methods can be constructed using a compass and straightedge. One may consider other methods for outlier detection such as Grubbs tests of Gauss pdfs [17] for one-sided and two-sided detection of outliers. This work is only the first step to using the described estimators of triangle distributions. In the future, we aim to address other issues such as the description of errors and uncertainties in measurements, estimation with small numbers of samples, and other rules for detecting and rejecting the outliers.

We conclude this section with a small Monte Carlo simulation study. We considered $n = 10, 20, 30, 40, 50$. For each n , we ran 100 simulations where we chose c at random in $(0, 1)$ and generated n numbers according to standard triangular distribution (using the Scientific Python library). We then estimated c by maximum likelihood and by the suggested method using $H(\mu)$ and $H(M)$. We computed the mean of the relative error for each n and summarized the results in the Table 5. As we can see, for large n the accuracy of the median-based method is always better than using the maximum Likelihood estimation. The proposed method is much simpler computationally.

5 Estimation of tail probabilities with deviations

We now turn to compute some tail probabilities. The quadratic form of the CDF $F(x)$ for the triangular distribution (Equation 1) results in a simple quadratic form of tail probabilities. For example, for any $\epsilon < \min\{c - a, b - c\}$ we have

$$P(|X - c| \geq \epsilon) = P(X \leq c - \epsilon) + P(X \geq c + \epsilon) = F(c - \epsilon) + (1 - F(c + \epsilon)) = \frac{((c-a)-\epsilon)^2}{(b-a)(c-a)} + \frac{((b-c)-\epsilon)^2}{(b-a)(b-c)} \quad (22)$$

The above probability can be constructed geometrically. Recall that one of our assumptions is having a segment of unit length. This allows us to compute both the products and ratios of segments.

Let us examine how the MAD deviations can be used to estimate tail probabilities. The most general bound for any distribution (with finite variance) is the Chebyshev’s inequality [18]:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (23)$$

This inequality is useful for $k \geq 1$. This inequality follows from the so-called Pearson inequality [19] with $r = 2$

$$P(|X - \mu| \geq kV_r^{1/r}) \leq \frac{1}{k^2}, \quad \text{where} \quad V_r = E(|X - \mu|^r) \quad (24)$$

For $r = 1$, a much less-known inequality exists for bounds in terms of mean absolute deviation $H(\mu)$ from the mean, namely

$$P(|X - \mu| \geq kH(\mu)) \leq \frac{1}{k} \quad (25)$$

TABLE 4 Comparison of outlier detection methods.

| Method | Parameters | L | $x_i \leq L$ | U | $x_i \geq U$ |
|------------------------------------|-------------------|------|--------------------|------|--------------------|
| Gauss | $\mu^* = 5.69$ | 3.37 | 3.2 | 8.01 | 8.2, 8.5, 8.6 |
| | $\sigma^* = 1.16$ | | | | |
| MLE | $\mu^* = 5.82$ | 3.40 | 3.2, 3.4 | 8.24 | 8.5, 8.6 |
| | $\sigma^* = 1.21$ | | | | |
| Mean-based | $\mu^* = 5.27$ | 3.05 | 3.2, 3.4 | 7.49 | 8.5, 8.6 |
| | $\sigma^* = 1.11$ | | | | |
| Mid-range | $\mu^* = 5.90$ | 3.70 | 3.2, 3.4, 3.5, 3.5 | 8.10 | 8.2, 8.5, 8.6 |
| | $\sigma^* = 1.10$ | | 3.6, 3.6 | | |
| MAD-based | $M^* = 5.80$ | 4.04 | 3.2, 3.4, 3.5, 3.5 | 7.56 | 7.9, 8.2, 8.5, 8.6 |
| | $H^* = 0.88$ | | 3.6, 3.6, 3.9 | | |
| Tukey's method (non-parametric) | $Q_1 = 5.80$ | 2.60 | None | 8.52 | 8.6 |
| | $Q_3 = 6.30$ | | | | |

TABLE 5 Relative errors in estimating mode c by Monte Carlo.

| Sample size n | $n = 10$ | $n = 20$ | $n = 30$ | $n = 40$ | $n = 50$ |
|-----------------|----------|----------|----------|----------|----------|
| MLE | 1.77 | 1.74 | 4.28 | 1.51 | 3.70 |
| Mean-based | 1.51 | 1.23 | 2.47 | 0.63 | 1.05 |
| MAD-based | 0.85 | 1.17 | 1.75 | 0.58 | 0.81 |

Similarly, there is an inequality in terms of the mean absolute deviation from the median $H(M)$ given by [9]

$$P(|X - M| \geq kH(M)) \leq \frac{1}{k} \tag{26}$$

Let us compute bounds based on the MAD deviation. First, define $\delta = H(\mu)/\sigma$. Note that for all distributions we have $H(\mu) \leq \sigma$ and therefore $\delta \leq 1$. We can re-write the MAD-based inequality for $H(\mu)$ in Equation (25) in terms of σ as follows

$$P(|X - \mu| \geq k\sigma) = P\left(|X - \mu| \geq \frac{k\sigma}{H(\mu)} \cdot H(\mu)\right) \leq \frac{\delta}{k} \tag{27}$$

Also, we can consider the Peek inequality [20]

$$P(|X - \mu| \geq k\sigma) \leq \frac{1 - \delta^2}{k^2 - 2k\delta + 1} \tag{28}$$

Comparing Equations (23) and (27) we find that MAD-based upper bound for $H(\mu)$ is lower than Chebyshev's upper bound for $1 \leq k \leq 1/\delta$. Comparing Equations (27) and (28) we find that Peek inequality is lower than both MAD and Chebyshev for $k > 1/\delta$.

We can now apply these bounds to triangular distribution. Since σ , $H(M)$, and $H(\mu)$ are shift-invariant and multiplicative, we can consider only the standard triangular distribution with $a = 0$ and $b = 1$. Consider again the symmetric case $c = 1/2$. From

Table 2, we have $\mu = 1/2$, $\sigma = \sqrt{6}/12$, $H(\mu) = 1/6$, and $\delta = \sqrt{6}/3$. The tail probability from Equation (1) is

$$P(|X - \mu| \geq k\sigma) = 1 + \frac{k^2}{6} - \frac{k\sqrt{6}}{3} \tag{29}$$

Let us now consider bounds using deviations. Therefore, from Equation (27), the MAD-based bound for the tail probability in standard symmetric triangular distribution is

$$P(|X - \mu| \geq k\sigma) \leq \frac{\sqrt{6}}{3k} \tag{30}$$

Comparing this with Chebyshev's inequality (23) we find that the MAD-based bound is sharper for $k \leq 1/\delta = \sqrt{6}/2 \approx 1.22$. The Peek inequality in Equation (28) (with $\delta = \sqrt{6}/3$) gives the following:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{3k^2 - 2k\sqrt{6} + 3} \tag{31}$$

This gives us a sharper bound than MAD for the same value $k = \sqrt{6}/2$.

In Figure 8 we compare the exact value of tail probabilities (Equation 29), the Chebyshev bound (Equation 23), the MAD bound (Equation 27), and the Peek bound (Equation 28). For $k < \sqrt{6}/2$ the MAD bound in Equation (27) gives a lower bound than either Chebyshev or Peek inequality. For $k > \sqrt{6}/2$, both Chebyshev and Peek give lower values with the Peek inequality being the sharper bound. We note that none of these bounds give good approximations, especially for larger k . For example, at point $k = \sqrt{6}/2$ the Chebyshev, MAD, and Peek bounds are $2/3$ whereas the exact value from Equation (29) is 0.25. We will consider other Chebyshev's type inequalities [21] and the inequalities based on the median in future work.

6 Geometry of quartile deviations

In quantile statistics, one often uses half of the interquartile range (IQR) as a measure of deviation [1]. If the first and third

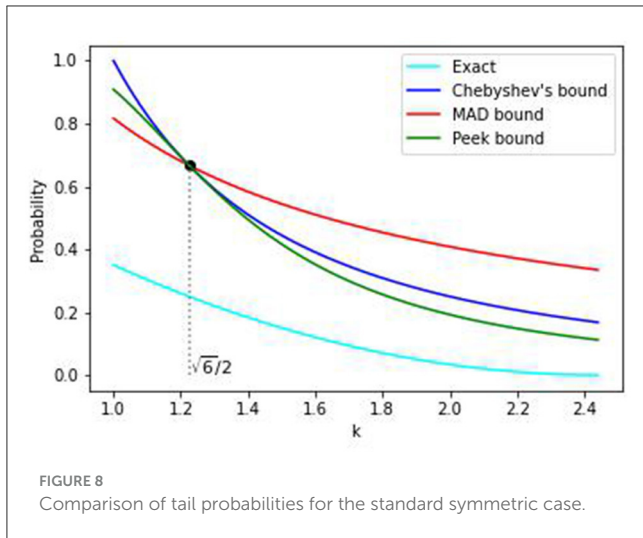


FIGURE 8 Comparison of tail probabilities for the standard symmetric case.

quartiles are given by $Q_1 = F^{-1}(1/4)$ and $Q_3 = F^{-1}(3/4)$ respectively then the interquartile range $I = (Q_3 - Q_1)$. The quartile deviation G is then $G = (Q_3 - Q_1)/2$. This has a very simple interpretation. Since $G = ((Q_3 - M) + (M - Q_1))/2$, we can interpret G as the average distance from the median to the other two quartiles.

To compute the quartiles and to derive the expression for G for triangular distribution, it will be enough to consider the case $a \leq c \leq (b - a)/2$. If $X \sim \text{Tr}(a, b, c)$ consider the random variable $Y = (a + b) - X$. This random variable has a triangular distribution $Y \sim \text{Tr}(a, b, (b + a) - c)$. Its quartiles and median are related to the quartiles and median of X by simple relationships

$$Q_1(X) - a = b - Q_3(Y), M(X) - a = b - M(Y), Q_3(X) - a = b - Q_1(Y) \tag{32}$$

For a geometric interpretation of the quartiles, we will find it convenient to find their distances from a or b . We compute the general expression for these distances by considering the following cases:

- Case 1 ($c = a$): For this case, the cumulative distribution function $F(x) = 1 - (b - x)^2/(b - a)^2$. The first quartile Q_1 , the median M and the third quartile are found from solving $F(x) = q$ for $q = 1/4, q = 1/2$ and $q = 3/4$ respectively. Since $x > c$ we use the second equation in Equation (3) and obtain:

$$(b - Q_1) = \frac{\sqrt{3}}{2}(b - a), (b - M) = \frac{\sqrt{2}}{2}(b - a), (b - Q_3) = \frac{(b - a)}{2}$$

This suggests the following geometric construction for quartiles:

1. construct $60^\circ/30^\circ$ right triangle T with hypotenuse $(b - a)$
2. compute the other two sides of T : $l_1 = (b - a)\sqrt{3}/2$ and $l_2 = (b - a)/2$
3. $l_1 = (Q_1 - b)$ and $l_2 = (b - Q_3)$ are distances of quartiles from b

4. the difference $(l_1 - l_2)$ is the interquartile range $I = (Q_3 - Q_1)$

- Case 2 ($0 < c \leq (a + b)/2$): from Equation (2.2) in Section 2 we have $M \geq c$. And since $Q_3 > M$ we obtain the following for the median M and the third quartile Q_3 :

$$(b - M) = \frac{\sqrt{2}}{2} \sqrt{(b - a)(b - c)}, (b - Q_3) = \frac{1}{2} \sqrt{(b - a)(b - c)}$$

What about the first quartile Q_1 ? If $Q_1 \leq c$ then from Equation (3) we have

$$a + \frac{1}{2} \sqrt{(b - a)(c - a)} \leq c \implies c \geq a + \frac{(b - a)}{4}$$

Similarly, for $c < a + (b - a)/4$ we have $Q_1 > c$. Therefore, from Equation (3) we obtain for Q_1 :

$$\begin{cases} (b - Q_1) = \frac{\sqrt{3}}{2} \sqrt{(b - a)(b - c)} & a < c < a + \frac{(b - a)}{4} \\ (Q_1 - a) = \frac{1}{2} \sqrt{(b - a)(c - a)} & a + \frac{(b - a)}{4} \leq c < \frac{(a + b)}{2} \end{cases}$$

This suggests the following geometrical construction for quartiles.

For the case $a < c < a + (b - a)/4$ we do the following:

1. compute the geometric mean $l_1 = \sqrt{(b - a)(b - c)}$
2. construct a $60^\circ/30^\circ$ right triangle T with hypotenuse l_1
3. compute the other two sides of this triangle: $l_2 = l_1\sqrt{3}/2$ and $l_3 = l_1/2$
4. $l_2 = (b - Q_1)$ and $l_3 = (b - Q_3)$ are distances from quartiles from b
5. the difference of sides $(l_1 - l_2)$ is the interquartile range $I = (Q_3 - Q_1)$

For the case $a + (b - a)/4 \leq c < (a + b)/2$ the construction is similar except that we use an additional $45^\circ/45^\circ$ triangle as follows:

1. compute the geometric mean $\sqrt{(b - a)(c - a)}$
2. bisect this geometric mean to obtain l_1
3. compute the geometric mean $\sqrt{(b - a)(b - c)}$
4. bisect this geometric mean to obtain l_2
5. $l_1 = (Q_1 - a)$ and $l_2 = (b - Q_3)$ are distances from quartiles from a and b respectively
6. the difference $(b - a) - (l_1 + l_2)$ is the interquartile range $I = (Q_3 - Q_1)$

- Case 3 ($(a + b)/2 \leq c < b$): In this case, from Equation (32) we obtain

$$(Q_1 - a) = \frac{1}{2} \sqrt{(b - a)(c - a)}, (M - a) = \frac{\sqrt{2}}{2} \sqrt{(b - a)(c - a)}$$

And for Q_3 , we can show (similar to the previous case)

$$\begin{cases} (b - Q_3) = \frac{1}{2} \sqrt{(b - a)(b - c)} & \frac{(a + b)}{2} < c < a + \frac{3(b - a)}{4} \\ (Q_3 - a) = \frac{\sqrt{3}}{2} \sqrt{(b - a)(c - a)} & a + \frac{3(b - a)}{4} \leq c < b \end{cases}$$

The geometric construction for this case is similar to Case 2.

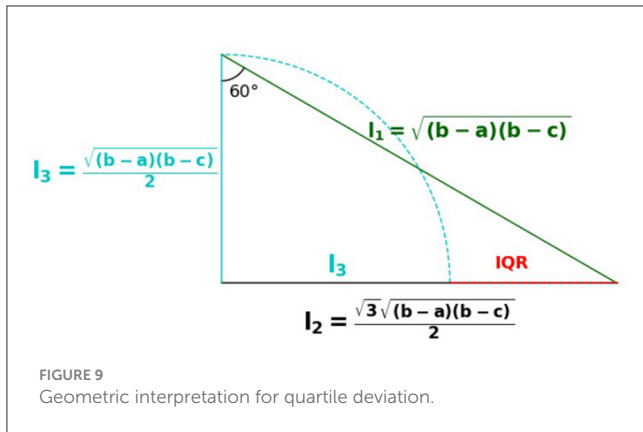


FIGURE 9 Geometric interpretation for quartile deviation.

- Case 4 ($c = b$): From Case $c = a$ and Equation (32) we obtain

$$(Q_1 - a) = \frac{1}{2}(b - a), (M - a) = \frac{\sqrt{2}}{2}(b - a), (Q_3 - a) = \frac{\sqrt{3}}{2}(b - a)$$

The geometric construction for this case is similar to Case 1.

Combining the expressions for the quartiles from the above cases, we can write a general expression for the quartile deviation:

$$G = \begin{cases} \frac{(\sqrt{3}-1)(b-a)}{4} & c = a \text{ or } c = b \\ \frac{(\sqrt{3}-1)\sqrt{(b-a)(c-a)}}{4} & a < c < a + \frac{(b-a)}{4} \\ \text{or } a + \frac{3(b-a)}{4} \leq c < b \\ \frac{(b-a)}{2} - \frac{\sqrt{b-a}}{4} (\sqrt{b-c} + \sqrt{c-a}) & a + \frac{(b-a)}{4} \leq c < a + \frac{3(b-a)}{4} \end{cases} \quad (33)$$

From the above, we have the following geometrical interpretation: the distances of the quartiles Q_1 and Q_3 can be computed as sides of right triangles. The difference in these distances is the interquartile range I . The quartile deviation G is one-half of this difference. The geometric interpretation of G can be demonstrated in Figure 9.

For the standard triangular distribution with $a = 0$ and $b = 1$, we can re-write the above as

$$G = \begin{cases} (\sqrt{3}-1)/4 & c = 0 \text{ or } c = 1 \\ (\sqrt{3}-1)\sqrt{c}/4 & 0 < c < 1/4 \text{ or } 3/4 \leq c < 1 \\ 1/2 - (\sqrt{1-c} + \sqrt{c})/4 & 1/4 \leq c < 3/4 \end{cases} \quad (34)$$

In Table 6 we compute the quartiles and the quartile deviation for the normalized triangular distribution with $a = 0$ and $b = 1$ for $c = 0, c = 1/4, c = 1/2, c = 3/4$ and $c = 1$.

7 Comparison of deviation measures

In this section, we present some results in comparison of the four deviation measures

1. standard deviation σ
2. mean absolute deviation (about mean) $H(\mu)$
3. mean absolute deviation (about median) $H(M)$
4. quartile deviation G

All these measures are shift-invariant. It is easy to show that scaling the data by any constant k , changes these deviation measures by $|k|$. Therefore, without loss of generality, it would be sufficient to compare these measures for the standard triangular distribution with $a = 0$ and $b = 1$.

In Table 7, we present a comparison for different values of c .

From our discussion in Section 7 we know that $H(M) \leq H(\mu) \leq \sigma$. We will first prove that $G < H(M)$. By symmetry, we will only need to prove $G < H(M)$ for $0 \leq c \leq 1/2$. We need to consider the following cases:

- Case 1: $c = 0$ or $c = 1$: From Equations (14) and (34) we have

$$H(M) - G = \frac{(2 - \sqrt{2})}{3} - \frac{(\sqrt{3} - 1)}{4} > 0$$

- Case 2: $0 < c \leq 1/4$: For this case, $(2 - c) \geq 7/4, \sqrt{c} \leq 1/2$, and $\sqrt{2(1-c)} \leq \sqrt{2}$. Therefore, from Equations (14) and (34) we have

$$H(M) - G > \frac{(7/4 - \sqrt{2})}{3} - \frac{(\sqrt{3} - 1)}{8} > 0$$

- Case 3: $1/4 \leq c \leq 1/2$: In this case, $(2 - c) \geq 3/2, \sqrt{c} \geq 1/2$, $\sqrt{1-c} \geq \sqrt{3}/2$, and $\sqrt{2(1-c)} \leq \sqrt{2}/2$. Therefore, from Equations (14) and (34) we have

$$H(M) - G \geq \frac{(3/2 - \sqrt{2}/2)}{3} - \frac{1}{2} + \frac{(\sqrt{3}/2 + 1/2)}{4} = \frac{(1 + \sqrt{3})}{8} - \frac{\sqrt{2}}{6} > 0$$

Therefore, we have the following relationship between these deviation measures:

$$G < H(M) \leq H(\mu) \leq \sigma \quad (35)$$

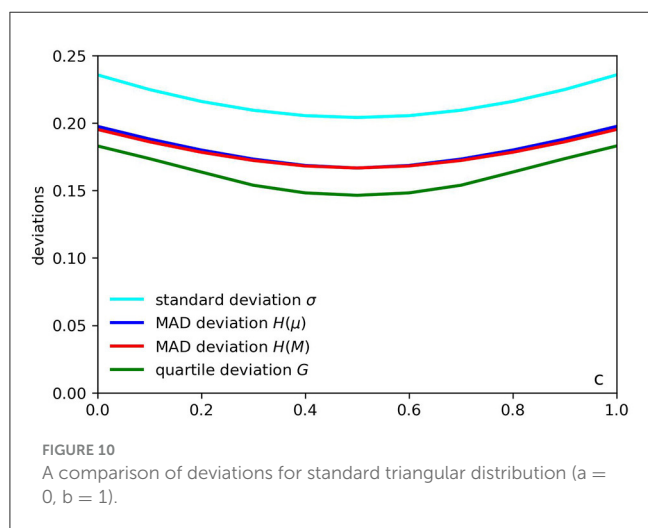
In Figure 10, we show the comparison between these four deviation measures for the standard triangular distribution. We note that across this range, both $H(\mu)$ and $H(M)$ are practically identical. At midpoint $c = 1/2$ they achieve their minimum value and are equal: $H(M) = H(\mu) = 1/6$, whereas at $c = 0$ or $c = 1$, they are within 2% of each other by relative value: $H(\mu) = (2/3)^4 \approx 0.198$ vs. $H(M) = (2 - \sqrt{2})/3 \approx 0.195$. The standard deviation varies from $\sigma = \sqrt{2}/6 \approx 0.236$ for $c = 0$ or $c = 1$. It achieves its minimum value $\sigma = \sqrt{6}/12 \approx 0.204$ at midpoint $c = 1/2$. Therefore, compared to mean absolute deviations $H(\mu)$ and $H(M)$, the standard deviation is σ is about 20% higher than the mean absolute deviation $H(M)$. If we consider the quartile deviation G , it has the maximum value $G = (\sqrt{3} - 1)/4 \approx 0.183$ for $c = 0$ and $c = 1$. For these values of $c, H(M) \approx 0.195$, and therefore, the quartile deviation is about 6% lower than $H(M)$. It decreases as c increases from $c = 0$ to $c = 1/2$ and achieves its minimum value at the midpoint: $G = (2 - \sqrt{2})/4 \approx 0.146$. Compared with the mean absolute deviation $H(M) \approx 0.167$ at the midpoint, the quartile

TABLE 6 Quartiles, quartile deviations and comparison of deviation measures for normalized triangular distributions.

| c | 0 | 0.25 | 0.5 | 0.75 | 1 |
|----------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Q ₁ | $\frac{2-\sqrt{3}}{2} \approx 0.13$ | 0.25 | $\frac{\sqrt{2}}{4} \approx 0.35$ | $\frac{\sqrt{3}}{4} \approx 0.43$ | 0.5 |
| M | $\frac{2-\sqrt{2}}{2} \approx 0.29$ | $\frac{4-\sqrt{6}}{4} \approx 0.39$ | 0.5 | $\frac{\sqrt{6}}{4} \approx 0.61$ | $\frac{\sqrt{2}}{2} \approx 0.71$ |
| Q ₃ | 0.5 | $\frac{4+\sqrt{3}}{4} \approx 0.57$ | $\frac{4+\sqrt{2}}{4} \approx 0.65$ | 0.75 | $\frac{\sqrt{3}}{2} \approx 0.87$ |
| G | $\frac{\sqrt{3}-1}{4} \approx 0.18$ | $\frac{3-\sqrt{3}}{8} \approx 0.16$ | $\frac{2-\sqrt{2}}{4} \approx 0.15$ | $\frac{3-\sqrt{3}}{8} \approx 0.16$ | $\frac{\sqrt{3}-1}{4} \approx 0.18$ |

TABLE 7 A comparison of deviation measures.

| c | 0 | 0.25 | 0.5 | 0.75 | 1 |
|----------|-------------------------------------|--------------------------------------------|-------------------------------------|--------------------------------------------|-------------------------------------|
| σ | $\frac{\sqrt{2}}{6} \approx 0.24$ | $\frac{\sqrt{26}}{24} \approx 0.21$ | $\frac{\sqrt{6}}{12} \approx 0.20$ | $\frac{\sqrt{26}}{24} \approx 0.21$ | $\frac{\sqrt{2}}{6} \approx 0.24$ |
| $H(\mu)$ | $(\frac{2}{3})^4 \approx 0.20$ | $\frac{1}{9} (\frac{7}{6})^3 \approx 0.18$ | $\frac{1}{6} \approx 0.17$ | $\frac{1}{9} (\frac{7}{6})^3 \approx 0.18$ | $(\frac{2}{3})^4 \approx 0.20$ |
| $H(M)$ | $\frac{2-\sqrt{2}}{3} \approx 0.20$ | $\frac{5-2\sqrt{2}}{12} \approx 0.18$ | $\frac{1}{6} \approx 0.17$ | $\frac{7-2\sqrt{6}}{12} \approx 0.18$ | $\frac{2-\sqrt{2}}{3} \approx 0.20$ |
| G | $\frac{\sqrt{3}-1}{4} \approx 0.18$ | $\frac{3-\sqrt{3}}{8} \approx 0.16$ | $\frac{2-\sqrt{2}}{4} \approx 0.15$ | $\frac{3-\sqrt{3}}{8} \approx 0.16$ | $\frac{\sqrt{3}-1}{4} \approx 0.18$ |



deviation is about 13% lower. We can conclude that, on average, the quartile deviation is about 10% lower than the mean absolute deviation (about median) $H(M)$. In contrast, the standard deviation is about 20% higher than $H(M)$. At the same time, the quartile deviation has higher volatility. We can compare these measures' volatilities $V(\cdot)$ by examining their range relative to their minimum value (at point $c = 1/2$). In percentage terms for the four deviation measures, we obtain:

$$V(\sigma) = 15.5\%, V(H(\mu)) = 18.5\%, V(H) = 17.2\%, V(G) = 25.0\%$$

The volatilities of standard and mean absolute deviations are close to each other by value and are around 15–18%. In contrast, the volatility of the quartile deviation is much higher at about 25%.

We should emphasize that, in practice, we have a limited number of points. As a result, it may be difficult to get accurate estimates for the quartiles to compute the quartile deviation. Therefore, from a practical standpoint, it is better to compute the mean absolute deviation and use it to compute the parameters of the distributions, as we have suggested.

8 Conclusion

In this paper, we presented a detailed comparison of deviation metrics for the the triangular distribution. We derived a novel formula for the mean absolute deviation from the median $H(M)$ and suggested a new estimation procedure of parameters. Most performance measures have simple geometric interpretation and can be constructed by compass and straightedge. Future work will consider applying these ideas to extensions of triangular distributions such as PERT and two-sided power distributions widely used in risk analysis.

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

EP: Writing—original draft, Writing—review & editing. YW: Writing—original draft, Writing—review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors thank metropolitan College of Boston University for their support.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Johnson NL, Kotz S. *Distributions in Statistics*. New York: J Wiley. (1970).
- Fairchild KW, Misra L, Shi Y. Using triangular distribution for business and finance simulations in excel. *J Financ Applic*. (2016) 42:313–36.
- Hopkinson M. Three point estimates: a brief guide, practical risk management series. *PM World J*. (2022) 6:1–3.
- Johnson D. Triangular approximations for continuous random variables in risk analysis. *J Oper Res Soc*. (2002) 53:457–67. doi: 10.1057/palgrave.jors.2601330
- Dorp JR, Kotz S. A novel extension of the triangular distribution and its parameter estimation. *Statistician*. (2002) 51:63–79. doi: 10.1111/1467-9884.00299
- Johnson D. The triangular distribution as a proxy for the beta distribution in risk analysis. *Statistician*. (1997) 46:387–98. doi: 10.1111/1467-9884.00091
- Johnson NL, Kotz S. Non-smooth sailing or triangular distributions revisited after some 50 years. *Statistician*. (1999) 48:179–87. doi: 10.1111/1467-9884.00180
- Gorard S. Revisiting a 90-year-old debate: the advantages of the mean deviation. *Br J Educ Stud*. (2005) 53:417–30. doi: 10.1111/j.1467-8527.2005.00304.x
- Pham-Gia T, Hung TL. The mean and median absolute deviations. *J Mathem Comput Model*. (2001) 34:921–36. doi: 10.1016/S0895-7177(01)00109-1
- Pinsky E, Klawansky S. MAD (about Median) vs. quantile-based alternatives for classical standard deviation, skew, and kurtosis. *Front Appl Mathem Statist*. (2023) 9:1206537. doi: 10.3389/fams.2023.1206537
- AbouRizk SM. *Input Modeling for Construction Simulation*. PhD Thesis, Purdue University Ann Arbor, MI: UMI Dissertation Services. (1990).
- Keefer DL, Bodily SE. Three-point approximations for continuous random variables. *Manage Sci*. (1983) 29:595–609. doi: 10.1287/mnsc.29.5.595
- Kotz S, Dorp JR. *Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications*. New Jersey: World Scientific Publishing. (2004). doi: 10.1142/5720
- Oliver EH. A maximum likelihood oddity. *Am Stat*. (1972) 26:43–4. doi: 10.1080/00031305.1972.10478930
- Friedman D, Diaconis P. On the histogram as a density estimator L_2 theory. *Probab Theory Relat Fields*. (1981) 57:453–76. doi: 10.1007/BF01025868
- Tukey J. *Exploratory Data Analysis*. London, UK: Pearson Publishing. (1977).
- Grubbs FE. Procedures for detecting outlying observations in samples. *Technometrics*. (1969) 11:1–21. doi: 10.1080/00401706.1969.10490657
- Feller JJ. *Probability Theory and Applications*. New York: J Wiley. (1956).
- Stuart A, Ord JK. *Kendalls Advanced Theory of Statistics*. vol 1 New York: Oxford University Press. (1987).
- Peek RL. Some new theorems on limits of variation. *Bull Am Mathem Soc*. (1933) 39:953–9. doi: 10.1090/S0002-9904-1933-05780-4
- Savage IR. Probability inequalities of the tchebycheff type. *J Res Nat Bureau Stand B Mathem Phys*. (1961) 65B:211–226. doi: 10.6028/jres.065B.020
- Bloomfield P, Steiger WL. *Least Absolute Deviations: Theory, Applications and Algorithms*. Boston: Birkhauser. (1983). doi: 10.1007/978-1-4684-8574-5
- Shad SM. On the minimum property of the first absolute moment. *Am Stat*. (1969) 23:27. doi: 10.2307/2682577
- 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. doi: 10.2307/2684955

Appendix

Appendix A: Details on computation of mean absolute deviations

We start by establishing a lower and upper bound for mean absolute deviations. Since $f(x) \geq 0$, is integrable and $E(X) < \infty$ we have $-|x - p|f(x) \leq (x - p)f(x) \leq |x - p|f(x)$ and, therefore, we obtain

$$H(p) = \int_X |x - p|f(x) dx \geq \left| \int_X (x - p)f(x) dx \right| = |p - \mu| \quad (36)$$

To establish an upper bound, we use the well-known fact that $H(M) \leq H(p)$ for any value of p [22–24]. In particular, if $p = \mu$ then $H(M) \leq H(\mu)$.

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 obtain an upper bound for $H(M)$:

$$\sigma^2 = E(g(|X - \mu|)) \geq g(E(|X - \mu|)) = [E(|X - \mu|)]^2 \geq H^2(M) \quad (37)$$

It follows from Equation (37) that $\sigma \geq H(\mu) \geq H(M)$.

We now deriving a formula for $H(p)$. We will find it convenient to define and evaluate the following auxiliary integrals $I_1(t)$ and $I_2(t)$. For $a \leq t \leq c$ define

$$I_1(t) = \int_t^c xf(x) dx = \int_t^c \frac{2x(x - a)}{(b - a)(c - a)} dx \quad (38)$$

$$= \frac{2c^3 - 2t^3 - 3ac^2 + 3at^2}{3(b - a)(c - a)}$$

and for $c \leq t \leq b$ define

$$I_2(t) = \int_c^t xf(x) dx = \int_c^t \frac{2x(b - x)}{(b - a)(b - c)} dx \quad (39)$$

$$= \frac{3bt^2 - 3bc^2 - 2t^3 + 2c^3}{3(b - a)(b - c)}$$

To derive a general expression for $H(p)$, we need to consider two cases: $p \geq c$ and $p \leq c$

Case 1: $p \geq c$. For this case, we have

$$H(p) = \int_a^c (p - x)f(x) dx + \int_c^p (p - x)f(x) dx + \int_p^b (x - p)f(x) dx$$

$$= p \left[\int_a^c f(x) dx - I_1(a) \right] +$$

$$p \left[\int_c^p f(x) dx - I_2(p) \right] + \left[I_2(b) - I_2(p) - p \int_p^b f(x) dx \right]$$

$$= I_2(b) - I_1(a) - 2I_2(p) + p(2F(p) - 1)$$

Case 2: $p \leq c$. For this case, we have

$$H(p) = \int_a^p (p - x)f(x) dx + \int_p^c (x - p)f(x) dx + \int_c^b (x - p)f(x) dx$$

$$= \left[p \int_a^p f(x) dx - (I_1(a) - I_1(p)) \right]$$

$$+ \left[I_1(p) - p \int_p^c f(x) dx \right]$$

$$+ \left[I_2(b) - p \int_c^b f(x) dx \right]$$

$$= I_2(b) - I_1(a) + 2I_1(p) + p(2F(p) - 1)$$

From these cases and noting that $I_1(a) + I_2(b) = \mu$, we obtain the following general formula for $H(p)$:

$$H(p) = \begin{cases} 2I_2(b) - \mu - 2I_2(p) + p(2F(p) - 1) & p \geq c \\ \mu - 2I_1(a) + 2I_1(p) + p(2F(p) - 1) & p \leq c \end{cases} \quad (40)$$

We can now compute the median absolute deviations $H(\mu)$ and $H(M)$

1. Computation of $H(\mu)$:

There are two cases to consider:

- Case 1: $a \leq c \leq (a + b)/2 \implies \mu \geq c$
- Case 2: $(a + b)/2 \leq c \leq b \implies \mu \leq c$

From this and Equation (40), we obtain the following general expression for the mean absolute deviation $H(\mu)$:

$$H(\mu) = \begin{cases} \frac{2(a+c-2b)^3}{81(a-b)(b-c)} & a \leq c \leq (a + b)/2 \\ \frac{2(b+c-2a)^3}{81(a-b)(a-c)} & (a + b)/2 \leq c \leq b \end{cases} \quad (41)$$

2. Computation of $H(M)$:

There are two cases to consider:

- Case 1: $a \leq c \leq (a + b)/2 \implies M > c$
- Case 2: $(a + b)/2 \leq c \leq b \implies M \leq c$

From this and Equations (4) and (40) we obtain the following general expression for $H(M)$:

$$H(M) = \begin{cases} \frac{(b-a)+(b-c)-\sqrt{2(b-a)(b-c)}}{3} & a \leq c \leq (a + b)/2 \\ \frac{(b-a)+(c-a)-\sqrt{2(b-a)(c-a)}}{3} & (a + b)/2 \leq c \leq b \end{cases} \quad (42)$$