 doi:10.20944/preprints201911.0268.v2

Statistical Mirroring: A Good Alternative Estimator of Dispersion

*Kabir Bindawa Abdullahi

Department of Biology, Faculty of Natural and Applied Sciences, Umaru Musa Yar'adua University, P.M.B., 2218 Katsina, Katsina State, Nigeria.

*Correspondence: kabir.abdullahi@umyu.edu.ng ; kabirnamallam@gmail.com
ORCID: 0000-0002-3810-9592

Abstract

The statistical properties of a good estimator include robustness, unbiasedness, efficiency, and consistency. However, the commonly used estimators of dispersion have lack or are weak in one or more of these properties. In this paper, I proposed statistical mirroring as a good alternative estimator of dispersion around defined location estimates or points. In the main part of the paper, attention is restricted to Gaussian distribution and only estimators of dispersion around the mean that functionalize with all the observations of a dataset were considered at this time. The different estimators were compared with the proposed estimators in terms of alternativeness, scale and sample size robustness, outlier biasedness, and efficiency. Monte Carlo simulation was used to generate artificial datasets for application. The proposed estimators (of statistical meanic mirroring) turn out to be suitable alternative estimators of dispersion that is less biased (more resistant) to contaminations, robust to scale and sample size, and more efficient to a random distribution of variable than the standard deviation, variance, and coefficient of variation. However, statistical meanic mirroring is not suitable with a mean (of a normal distribution) close to zero, and on a scale below ratio level.

Keywords: isomorphic optinalysis; dispersion; statistical mirrors; estimators; statistical properties

1. Introduction

A good estimator is statistically characterized as robust (invariant) to outliers and scale, unbiased, efficient to random samples, and consistent to sample sizes. However, the commonly used estimators of dispersion (e.g., standard deviation, variance, coefficient of variation, etc) lack one or more of these properties. Most of the robust estimators (e.g., interquartile range, quartile coefficient of dispersion, Qn and Sn by Rousseeuw-Croux, and other M-estimators) have low efficiency [1]; [2], while most of the relatively efficient estimators (e.g., standard deviation, variance, coefficient of variation) are not robust [1]. In practice, inefficient and inconsistent estimators could lead to a strong bias and erroneous statistical conclusions about a sample or population. Therefore, users have to be very cautious about which estimator is suitable and precise for their data, otherwise wrong statistical inference and conclusion would be drawn.

Under asymmetric distribution with some outliers, mean and standard deviation estimates lack robustness and lead to a strong bias [1], and the same thing would be expected from its derivative functions such as the coefficient of variation. Median and interquartile range are robust but inefficient estimators under contaminated and asymmetric distributions. In survey statistics, outliers are unavoidable elements, and their analysis must be in non-parametric form, otherwise, the data should be checked and corrected, or transformed to suit the parametric designs [1]. Unfortunately, data transformation (intended to remove, smooth, or normalize the existing outliers) is considered, in some cases, a dishonest and bias treatment [1].

In the comparison (of dispersion) of two or more disparate groups (that have means of very different magnitudes) and characteristics (that use different units of measurement), estimators such as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

variance, standard deviation, dispersion index (or variance-to-mean ratio), interdecile range, median absolute deviation from median (MADM), mean absolute deviation from mean (MAD), interquartile range (IQR), quartile coefficient of dispersion (derived from IQR), Qn and Sn by Rousseeuw-Croux, and other M-estimators have failed. The coefficient of variation stands as one of the important scale invariants (a standardized, unitless and dimensionless) estimator that have been used to compare datasets on different scales. The classic version of the coefficient of variation has been used in different areas such as biology [3], biochemistry [4], medical physics [5], neuroscience [6], [7], engineering [8], psychology [9], [10], sociology and economics [11]. Despite that, coefficient of variation is not always a good measure of relative dispersion and has the following pitfalls and drawbacks: a) it has no clear bounds that are within a fixed range b) inappropriate in asymmetric distribution, c) inappropriate with nominal, ordinal and interval scales d) very sensitive to outliers especially if the mean of the distribution tends close to zero, e) appropriately works with positive values of observations, f) inappropriate for comparing groups with different sample sizes [11].

Scale statistics (a measure of dispersion, scale, and spread of data) is an important primary stage for inferential statistics. Therefore, the goodness and statistical qualities of inferential statistics rely on the goodness of its scale estimators. The strict barriers between parametric and non-parametric statistics are importantly the normality assumption and homoscedasticity condition which depend on the distribution's shape and the scale invariance respectively. Therefore, having a robust estimator (i.e., robustness to outliers, scale, and sample size) and efficient breaks this barrier.

Recently, scientists, statisticians, and data analysts suffer on the choice of estimators (navigating from descriptive to inferential methods) that resist outliers and the underlying distribution of the data, and at the same time maintain its robustness, efficiency, and consistency. Among the good properties, users should look at for good estimators of dispersion included the following: robustness to scale and contaminations, unbiasedness, efficiency, and consistency; both at symmetric and asymmetric distributions. In this paper, a statistical mirroring is proposed as a good alternative estimator of dispersion that measures the coefficient of proximity, proximity, and deviation of all the data points of a variable about defined location estimates or points. I restricted attention to the use of Kabirian-based isomorphic optinalysis [12] to derive the concept of statistical mirroring. In application, attention is focused on Gaussian distribution and only estimators of dispersion around the mean that functionalize with all the observations of a dataset were considered at this time. In due course, absolute measures of dispersion, the variance and standard deviation; and a relative measure of dispersion, the coefficient of variation were used as the reference standard estimators of dispersion.

2. Preliminaries

Definition I. Optinalysis is a function that autoreflectively or isoreflectively compares the symmetry, similarity, and identity between two mathematical structures as a mirror-like (optic-like) reflection of each other about a symmetrical line or mid-point [12].

Definition II. Optinalysis is a function that is comprised of an assigned optical (mirror) scale (R) that bijectively re-maps ($\twoheadrightarrow$) isoreflective or autoreflective pair of mathematical structures. Figure 1 illustrates how isoreflective pairs of points are mapped and also re-mapped by an optical scale [12].

Definition III: Isoreflective Pair: Isoreflective pair refers to any two mathematical structures under reflection about a central midpoint. An isoreflectivity refers to the logical and meaningful essence or property of being isoreflective [12]

Definition IV. Isomorphic (Comparative) Optinalysis. Isomorphic optinalysis refers to the analysis (of similarity and identity measures) between isoreflective pair of mathematical structures under optinalysis. It is a method of similarity and identity measures [12].

 doi:10.20944/preprints201911.0268.v2

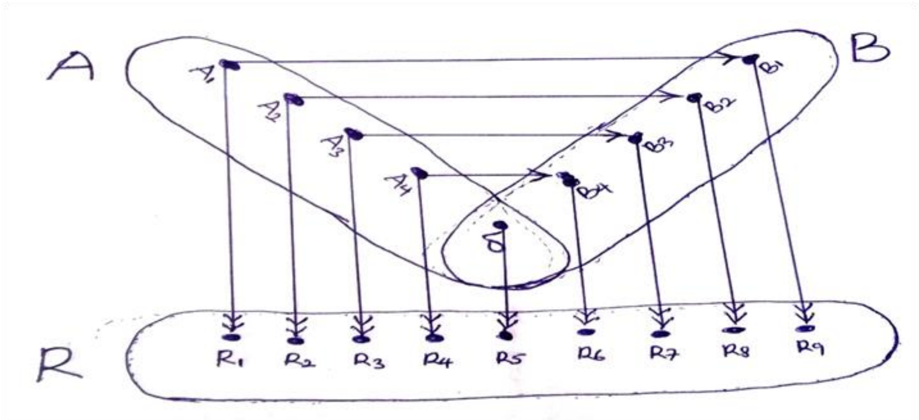

Figure 1: Mapping between isoreflective pair of points and re-mapping with the optical scale. A is a domain; B is a co-domain of A; δ is a mid-point or symmetrical line, and R is the optical scale. The symbol ⤀ indicates a *bijective mapping* between the isoreflective pair around a midpoint and ↠ indicates a *bijective re-mapping* by the optical scale R.

Definition V In comparative optinalysis, a reflection (pairing) is said to be head-to-head if the lower order elements (observations) of the isoreflective pair (of two mathematical structures) are extreme away from the midpoint [12].

$$A = (A_1, A_2, A_3, \dots \dots, A_n) \overset{\delta}{⤀} B = (B_n, \dots \dots, B_3, B_2, B_1)$$

Definition VI. In comparative optinalysis, a reflection or pairing is said to be tail-to-tail if the lower order elements (observations) of the isoreflective pair (of two mathematical structures) are extreme towards the midpoint [12].

$$A = (A_n, \dots \dots, A_3, A_2, A_1) \overset{\delta}{⤀} B = (B_1, B_2, B_3, \dots \dots, B_n)$$

Theorem I. Isoreflective pair of mathematical structures under optinalysis are similar and identical to a certain magnitude by a coefficient, called optinalytic coefficient (e.g, Kabirian coefficient, denoted as KC) [12].

Suppose we have an optinalytic construction of isoreflective pair with an assigned optical scale (R) as follows:

$$f: \begin{bmatrix} A = (A_1, A_2, A_3, \dots A_n) & \overset{\delta}{⤀} & B = (B_n, \dots, B_3, B_2, B_1) \\ \downarrow & \downarrow & \downarrow \\ R = (R_1, R_2, R_3, \dots R_n) & R_{n+1} & (R_{n+2}, \dots, R_{2n-1}, R_{2n}, R_{2n+1}) \end{bmatrix}$$

Such that: $(A_1, A_2, A_3, \dots, A_n) \in A$; $(B_1, B_2, B_3, \dots, B_n) \in B$; $\delta \notin A \,\&\, B$; $A, B \,\&\, R \in \mathbb{R}$; $R_1 \neq 0$; and $A \,\&\, B$ are isoreflective pair on a chosen pairing about a central line (δ). $\delta = 0$ is always the default operation, except if the function is under optinalytic normalization.

Then, the Kabirian coefficient of identity and similarity between the isoreflective pair is expressed as (Equations 1):

$$KC_{Sim./Id.}(A, B) = \frac{R_{n+1}(A_1 + A_2 + A_3 + \dots + A_n + \delta + B_n + \dots + B_3 + B_2 + B_1)}{(R_1.A_1) + (R_2.A_2) + (R_3.A_3) + \dots + (R_n.A_n) + (R_{n+1}.\delta) + (R_{n+2}.B_n) + \dots + (R_{2n-1}.B_3) + (R_{2n}.B_2) + (R_{2n+1}.B_1)} \quad (1)$$

$$\begin{cases} KC_{Sim./Id.}(A,B) = 1, & if\ |E(A)| = |E(B)| \\ KC_{Sim./Id.}(A,B) = 0, & if\ |E(A)| = -|E(B)|; -|E(A)| = |E(B)| \\ 0 \leq\ KC_{Sim./Id.}(A,B) \leq 1, & if\ |E(A)| < |E(B)| \\ 1 \leq KC_{Sim./Id.}(A,B) \leq n+1, & if\ |E(A)| > |E(B)| \\ KC_{Sim./Id.}(A,B) \geq n+1, < 0, & if\ |E(A)| > E|E(B)| \end{cases}$$

Where $|E(A)|$ and $|E(B)|$ are the absolute optical moment of A and B respectively about the central mid-point through a distance D started from the centre. It is expressed by equations (2.1) and (2.2).

$$f: \begin{bmatrix} A = (A_1, A_2, A_3, \dots A_n) & \overset{\delta}{\twoheadrightarrow} & B = (B_n, \dots, B_3, B_2, B_1) \\ \downarrow & \downarrow & \downarrow \\ D = (D_n, D_{n-1}, D_{n-2}, \dots D_1) & D_0 & (D_1, \dots, D_{n-2}, D_{n-1}, D_n) \end{bmatrix}$$

$$|E(A)| = |(D_n.A_1) + (D_{n-1}.A_2) + (D_{n-2}.A_3) + \cdots + (D_1.A_n)| = \left|\sum_{n=1}^{n}(D.A)\right| \tag{2.1}$$

$$|E(B)| = |(D_n.B_1) + (D_{n-1}.B_2) + (D_{n-2}.B_3) + \cdots + (D_1.B_n)| = \left|\sum_{n=1}^{n}(D.A)\right| \tag{2.2}$$

The Kabirian coefficient of symmetry, similarity, and identity, $KC_{Sim./Id.}(A,B)$ are translated to the proportion of significance, $P_{Sim./Id.}(A,B)$, using the equations (3) and (4).

$$P_{Sim./Id.}(A,B) = \frac{(nr_1 + r_1) - K_c(2nr_1 + r_1)}{K_c - (nr_1 + r_1)}, \forall\ 0 \leq K_c \leq 1 \tag{3}$$

$$\begin{cases} 0 \leq P_{Sim./Id.}(A,B) \leq 1, & if\ \frac{n+1}{2n+1} \leq\ KC_{Sim./Id.}(A,B) \leq 1 \\ -1 \leq P_{Sim./Id.}(A,B) \leq 0, & if\ 0 \leq\ KC_{Sim./Id.}(A,B) \leq \frac{n+1}{2n+1} \end{cases}$$

Or otherwise,

$$P_{Sim./Id.}(A,B) = \frac{(nr_1 + r_1) - r_1K_c}{(2nr_1 + r_1)K_c - (nr_1 + r_1)}, \forall\ 1 \leq K_c \leq n+1; K_c \geq n+1\ \ \&\ \forall\ K_c \leq 0 \tag{4}$$

$$\begin{cases} 0 \leq P_{Sim./Id.}(A,B) \leq 1, & if\ 1 \leq\ KC_{Sim./Id.}(A,B) \leq n+1 \\ -1 \leq P_{Sim./Id.}(A,B) \leq 0, & if\ KC_{Sim./Id.}(A,B) \geq n+1, \leq 0 \end{cases}$$

Subsequently, the proportion of asymmetry, dissimilarity, and none-identity, $P_{Asym./Dsim./Nid.}(A,B)$ are calculated from these using equations (5) and (6). Translation of Kabirian coefficient is valid if and only if the outcomes are within the range of values -1 to 1 (or -100 to 100 of its equivalent percentage).

$$\text{If } P_{Sim./Id.}(A,B) \geq 0, \text{ then } P_{Dsim./Nid.}(A,B) = 1 - P_{Sim./Id.}(A,B) \tag{5}$$

$$\text{If } P_{Sim./Id.}(A,B) \leq 0, \text{ then } P_{Dsim./Nid.}(A,B) = -1 - P_{Sim./Id.}(A,B) \tag{6}$$

3. Statistical Mirroring

3.1 Proposition:

The dispersion (proximity and deviation) of data points from a defined location of a given distribution is the isoreflectivity of its data points to a defined statistical mirror (i.e., a defined and amplified location estimate of the distribution).

Prove:

Suppose we have an ordered spread of a random variable $X = (x_1 \leq, x_2, \leq\ x_3, \leq \cdots, \leq x_n)$ or alternatively $X = (x_1 \geq, x_2, \geq x_3, \geq \cdots, \geq x_n)$ and its statistical mirror $P = (p_1, p_2, p_3, \ldots, p_n)$, then their optinalytic construction with an assigned optical scale $R = (r_1, r_2, r_3, \ldots, r_n)$ is expressed as follows:

$$f\colon X_n \overset{\delta}{\not\twoheadrightarrow} P_n \twoheadrightarrow R_n$$

$$f\colon X = (x_1 \leq, x_2, \leq x_3, \leq \cdots, \leq x_n) \overset{\delta}{\not\twoheadrightarrow} P = (p_1, p_2, p_3, \ldots, p_n) \twoheadrightarrow R = (r_1, r_2, r_3, \ldots, r_{2n+1})$$

$$f\colon \begin{bmatrix} X = (x_1 \leq, x_2, \leq x_3, \leq \cdots, \leq x_n) & \overset{\delta}{\not\twoheadrightarrow} & P = (p_1, p_2, p_3, \ldots p_n) \\ \downarrow & \downarrow & \downarrow \\ R = (r_1, r_2, r_3, \ldots, r_n) & r_{n+1} & (r_{n+2}, r_{n+3}, r_{n+4}, \ldots, r_{2n+1}) \end{bmatrix}$$

Or in alternative way,

$$f\colon X = (x_1 \geq, x_2, \geq x_3, \geq \cdots, \geq x_n) \overset{\delta}{\not\twoheadrightarrow} P = (p_1, p_2, p_3, \ldots, p_n) \twoheadrightarrow R = (r_1, r_2, r_3, \ldots, r_{2n+1})$$

$$f\colon \begin{bmatrix} X = (x_1 \geq, x_2, \geq x_3, \geq \cdots, \geq x_n) & \overset{\delta}{\not\twoheadrightarrow} & P = (p_1, p_2, p_3, \ldots, p_n) \\ \downarrow & \downarrow & \downarrow \\ R = (r_1, r_2, r_3, \ldots, r_n) & r_{n+1} & (r_{n+2}, r_{n+3}, r_{n+4}, \ldots, r_{2n+1}) \end{bmatrix}$$

Such that: $(x_1, x_2, x_3, \ldots, x_n) \in X$; $(p_1, p_2, p_3, \ldots p_n) \in P$; $\delta \notin X \,\&\, P$; $X, P \,\&\, R \in \mathbb{R}$; $R_1 \neq 0$; and $X \,\&\, P$ are isoreflective pairs. $\delta = 0$ is always the default operation, except if the function is under optinalytic normalization. P could be the average (mean), median, mode, maximum, minimum and etc of the distribution.

By Kabiran-based optinalysis, the Kabirian coefficient of proximity $(KC_{Prox.})$ of the data points about a defined location is expressed by equation (7):

$$KC_{Prox.}(X, P) = \frac{r_{n+1}(x_1 + x_2 + x_3 + \cdots + x_n + \ \delta + p_1 + p_2 + p_3 + \cdots p_n)}{(r_1.x_1) + (r_2.x_2) + (r_3.x_3) + \cdots + (r_n.x_n) + (r_{n+1}.\delta) + (r_{n+2}.p_1) + (r_{n+3}.p_2) + (r_{n+4}.p_3) + \cdots (r_{2n+1}.p_n)} \tag{7}$$

Since P_n is a constant factor to the arithmetic sum of R_n. Then we can shorten or simplify the equation (7) as equation (8):

$$KC_{Prox.}(X, P) = \frac{r_{n+1}(x_1 + x_2 + x_3 + \cdots + x_n + \ \delta + np)}{(r_1.x_1) + (r_2.x_2) + (r_3.x_3) + \cdots + (r_n.x_n) + (r_{n+1}.\delta) + p\left\{\frac{n}{2}[2r_{n+2} + (n-1)\Delta r]\right\}} \tag{8}$$

$$\begin{cases} KC_{Prox.}(X,P) = 1, & if\ |E(X)| = |E(P)| \\ KC_{Prox.}(X,P) = 0, & if\ |E(X)| = -|E(P)|;\ -|E(X)| = |E(P)| \\ 0 \leq\ KC_{Prox.}(X,P) \leq 1, & if\ |E(X)| < |E(P)| \\ 1 \leq KC_{Prox.}(X,P) \leq n+1, & if\ |E(X)| > |E(P)| \\ KC_{Prox.}(X,P) \geq n+1, < 0, & if\ |E(X)| > |E(P)| \end{cases}$$

Where $|E(X)|$ and $|E(P)|$ are the absolute optical moment of X and P respectively about the mid-point through a distance d started from the centre. It is expressed by equations (9.1) and (9.2).

$$f\colon \begin{bmatrix} X = (x_1, x_2, x_3, \dots x_n) & \overset{\delta}{\nrightarrow} & P = (p_n, \dots, p_3, p_2, p_1) \\ \downarrow & \downarrow & \downarrow \\ D = (d_n, d_{n-1}, d_{n-2}, \dots d_1) & d_0 & (d_1, \dots, d_{n-2}, d_{n-1}, d_n) \end{bmatrix}$$

$$|E(X)| = |(d_n.x_1) + (d_{n-1}.x_2) + (d_{n-2}.x_3) + \cdots + (d_1.x_n)| = \left|\sum_{n=1}^{n} (r_n.x_n)\right| \qquad (9.1)$$

$$|E(P)| = |(d_n.p_1) + (d_{n-1}.p_2) + (d_{n-2}.p_3) + \cdots + (d_1.p_n)| = \left|\sum_{n=1}^{n} (r_n.p_n)\right| \qquad (9.2)$$

The Kabirian coefficient of proximity, $KC_{Prox.}(X,P)$ are translated to a proportionate level of significance of proximity of data points about a defined location, $P_{Prox.}(X,P)$, using the equations (10) and (11).

$$P_{Prox.}(X,P) = \frac{(nr_1 + r_1) - K_c(2nr_1 + r_1)}{K_c - (nr_1 + r_1)}, \forall\ 0 \leq K_c \leq 1 \qquad (10)$$

$$\begin{cases} 0 \leq\ P_{Prox.}(X,P) \leq 1, & if\ \frac{2}{3} \leq\ KC_{Prox.}(X,P) \leq 1 \\ -1 \geq\ P_{Prox.}(X,P) \leq 0, & if\ 0 \leq\ KC_{Prox.}(X,P) \leq \frac{2}{3} \end{cases}$$

Or otherwise,

$$P_{Prox.}(X,P) = \frac{(nr_1 + r_1) - r_1 K_c}{(2nr_1 + r_1)K_c - (nr_1 + r_1)}, \forall\ 1 \leq K_c \leq n+1; K_c \geq n+1\ \&\ \forall\ K_c \leq 0 \quad (11)$$

$$\begin{cases} 0 \leq\ P_{Prox.}(X,P) \leq 1, & if\ 1 \geq\ KC_{Prox.}(X,P) \leq n+1 \\ -1 \leq\ P_{Prox.}(X,P) \leq 0, & if\ KC_{Prox.}(X,P) \geq n+1, \leq 0 \end{cases}$$

Subsequently, the proportionate significance level of deviation of data points about a defined location, $P_{Dev.}(X,P)$ is calculated from these using equations (12) and (13).

Translation of Kabirian coefficient is valid if and only if the outcomes are within the range of values -1 to 1 (or -100 to 100 of its equivalent percentage).

$$\text{If } P_{Prox.}(X,P) \geq 0, \text{ then } P_{Dev.}(X,P) = 1 - P_{Prox.}(X,P) \qquad (12)$$

$$\text{If } P_{Prox.}(X,P) \leq 0, \text{ then } P_{Dev.}(X,P) = -1 - P_{Prox.}(X,P) \qquad (13)$$

3.2 Different approaches of statistical mirroring

Suppose that $f\colon X_n \overset{\delta}{\nrightarrow} P_n \twoheadrightarrow R_n$, where X_n is the observations, P_n is the statistical mirror, R_n is the optinalytic scale. It is called:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

a. *Statistical meanic mirroring*, if and only if $P_n = M_n$ (the mean of the distribution of X_n). It measures the deviation or proximity (how far or close) of observations (data points) from its mean estimate.

b. *Statistical medianic mirroring*, if and only if $P_n = M_d$ (the median of the distribution of X_n). It measures the deviation or proximity (how far or close) of observations (data points) from its median estimate.

c. *Statistical modalic mirroring*, if and only if $P_n = M_o$ (the mode of the distribution of X_n). It measures the deviation or proximity (how far or close) of observations (data points) from its mode estimate.

d. *Statistical maximalic mirroring*, if and only if $P_n = M_x$ (the maximum of the distribution of X_n). It measures the deviation or proximity (how far or close) of observations (data points) from its maximum estimate.

e. *Statistical minimalic mirroring*, if and only if $P_n = M_y$ (the minimum of the distribution of X_n). It measures the deviation or proximity (how far or close) of observations (data points) from its minimum estimate.

f. *Statistical rangic mirroring*, if and only if $P_n = M_{x-y}$ (the range of the distribution of X_n). It measures the deviation or proximity (how far or close) of observations (data points) from its range estimate.

g. *Statistical reference mirroring*, if and only if $P_n = R_f$ (a reference value outside the distribution of X_n). It measures the deviation or proximity (how far or close) of observations (data points) from its reference estimate value.

h. *Endo-statistical mirroring*, if and only if $P_n = M_n, M_d, M_o, M_x, M_y, M_{x-y}$ (of a location estimate).

i. *Exo-statistical mirroring*, if and only if $P_n \neq M_n, M_d, M_o, M_x, M_y, M_{x-y}$ (of a location estimate).

3.3 Properties of statistical mirroring

i. It is based on all observations of the dataset. Therefore, extreme minimum and maximum values are not discarded.

ii. It applies to all sets of real numbers (such as discrete or continuous variables containing either or both negative and positive values),

iii. It may involve the use of mean and other defined location estimates such as median, maximum, minimum, and range.

iv. It is variant concerning changes in a location parameter. Therefore, it is not suitable for the comparison of multiple datasets (measurements) below the ratio level of scale.

$$KC_{Prox.}(x,p) \neq KC_{Prox.}\,(a+x, a+p) \neq KC_{Prox.}(-a+x, -a+p)$$

$$P_{Prox.}(x,p) \neq P_{Prox.}\,(a+x, a+p) \neq P_{Prox.}(-a+x, -a+p)$$

$$P_{Dev.}(x,p) \neq P_{Dev.}\,(a+x, a+p) \neq P_{Dev.}(-a+x, -a+p)$$

where $x, p, a = \mathbb{R}$

v. Statistical mirroring on ratio level of scale is scale-invariant (i.e., robust to scale, unitless and dimensionless) estimator of dispersion (scale). This property corresponds to the invariance of isomorphic optinalysis under translation transformation [12]. Therefore, it is very suitable with measurements on a ratio scale.

$$KC_{Prox.}(x,p) = KC_{Prox.}\,(ax, ap) = KC_{Prox.}(-ax, -ap)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

$$P_{Prox.}(x,p) = P_{Prox.}(ax,ap) = P_{Prox.}(-ax,-ap)$$

$$P_{Dev.}(x,p) = P_{Dev.}(ax,ap) = P_{Dev.}(-ax,-ap)$$

where $x, p, a = \mathbb{R}; a \neq 0$.

vi. Statistical mirroring is bi-coefficients and translative (i.e., forward and reverse translations). It gives two possible coefficients ($KC1_{Prox.}$, $KC2_{Prox.}$) due to its commutive property, but each coefficient translates into the same results ($P_{Prox.}$, and $P_{Dev.}$), which can be used to compute back to the two coefficients.

$$\begin{matrix} KC_{Prox.}(x,p) \\ \dots \\ KC_{Prox.}(p,x) \end{matrix} \rightleftharpoons P_{Prox.}(x,p) = P_{Prox.}(p,x) \rightleftharpoons P_{Dev.}(x,p) = P_{Dev.}(p,x)$$

The two possible Kabirian bi-coefficients work on two different optinalytic scales.

vii. Statistical mirroring is commutive around a central mid-point of the two isoreflective pairs. This property corresponds to the invariance of isomorphic optinalysis under central rotation (alternate reflection) transformation [12].

$$KC_{Prox.}(x,p) \rightleftharpoons KC_{Prox.}(p,x)$$

$$P_{Prox.}(x,p) = P_{Prox.}(p,x)$$

$$P_{Dev.}(x,p) = P_{Dev.}(p,x)$$

viii. Statistical mirroring is population independent and invariance to sample size. But the sample size invariance is effective to $P_{Prox.}$ and $P_{Dev.}$, and not to $KC_{Prox.}$.

Supposed we have the following set of observations, x_a, x_b:

$$x_a = (x)$$

$$x_b = (x, x, x, \dots, x_n)$$

And their statistical mirror as $p = (p_1, p_2, p_3, \dots, p_n)$, where p is the set of amplified mean estimates of x_a or x_b.

By Kabirian-based optinalysis, it shows that

$$KC_{MnProx.}(x_a, p) \neq KC_{MnProx.}(x_b, p)$$

$$P_{MnProx.}(x_a, p) = P_{MnProx.}(x_b, p)$$

$$P_{MnDev.}(x_a, p) = P_{MnDev.}(x_b, p)$$

ix. Statistical meanic mirroring only, not any others, is invariant to sample size or multiple repeats in the same order of all the observations of a dataset.

Supposed we have the following set of observations, x_a, x_b, and x_c:

$$x_a = (x^2, x^3, x^5, x^6, x^7, x^7, x^9)$$

$$x_b = (x^2, x^2, x^3, x^3, x^5, x^5, x^6, x^6, x^7, x^7, x^7, x^7, x^9, x^9)$$

$$x_c = (x^2, x^2, x^2, x^3, x^3, x^3, x^5, x^5, x^5, x^6, x^6, x^6, x^7, x^7, x^7, x^7, x^7, x^7, x^9, x^9, x^9)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

And their statistical meanic mirror as $p = (p_1, p_2, p_3, \ldots, p_n)$, where p is the set of amplified mean estimates of x_a, x_b, or x_c.

By Kabirian-based optinalysis, it shows that

$$KC_{MnProx.}(x_a, p) \neq KC_{MnProx.}(x_b, p) \neq KC_{MnProx.}(x_c, p)$$

$$P_{MnProx.}(x_a, p) = P_{MnProx.}(x_b, p) = P_{MnProx.}(x_c, p)$$

$$P_{MnDev.}(x_a, p) = P_{MnDev.}(x_b, p) = P_{MnDev.}(x_c, p)$$

4. Application and Methods Comparison on Dispersion Measures

In this paper, attention is restricted to symmetric distribution and only estimators that functionalize with all the observations of the datasets were considered. I apply the statistical mirroring (specifically the statistical meanic mirroring) to show its suitability as an alternative approach of dispersion measure around the mean. The proposed estimators (i.e., the statistical meanic mirroring) were compared based on desirable properties a good estimator should have, with the most used reference standard estimators of absolute and relative dispersion around the mean (i.e, standard deviation, variance, and coefficient of variation).

4.1 Artificial Datasets

Monte Carlos simulation was used to generate artificial datasets. A total of 1000 random variables were generated from a normal distribution with the following parameters: $\mu = 10;\ \sigma = 1;\ n = 10, 50, 100, 200, \&\ 500$. These steps of datasets generation was repeated with $\sigma = 2, 3, \ldots., 15$. This made a total of 75,000 parametrized random numbers.

Before outlier biasedness evaluation, Monte Carlos simulation was also used to generate 1000 random variables from a normal distribution with $\mu = 10;\ \sigma = 2;\ n = 50$. Then, a single point and 20% contaminations with a magnitude of contaminants of ±5, ±10, ±15, ±20, ±500, ±1000, ±5000, ±10000 were added to the upper and the lower values of the sorted distribution. Table 1 presented how the average mean of the distribution changed with the added contaminations. From table 1, the procedural design allows us to check the behaviors of the estimators close and away from a zero mean of the contaminated normal distribution.

Table 1: Change in the average mean due to contaminations. Distribution $[N\ (\mu = 10, \sigma = 2\ \&\ 1000\ random\ variables)]$.

Contaminations	Single Point PC	Single Point NC	20% PC	20% NC
±0	10.0044	10.0044	10.0044	10.0044
±5	10.1044	9.9044	11.0044	9.0044
±10	10.2044	9.8044	12.0044	8.0044
±15	10.3044	9.7044	13.0044	7.0044
±20	10.4044	9.6044	14.0044	6.0044
±500	20.0044	0.0044	110.0044	-89.9956
±1000	30.0044	-9.9956	210.0044	-189.9956
±5000	110.0044	-89.9956	1010.0044	-989.9956
±10000	210.0044	-189.9956	2010.0044	-1989.9956

Keys: PC = Positive contamination; NC = Negative contamination.

4.2 Data Analysis

Microsoft Excel statistical functions (Refer to supplementary files S1-S5) were used to estimate the Kabirian coefficient of meanic proximity, meanic proximity, meanic deviation, standard deviation, coefficient of variation and variance of the generated random variables. The two possible Kabirian bi-coefficients ($KC1_{Prox.}(x, p)$ and $KC2_{Prox.}(x, p)$) were evaluated and sorted on a standardized optinalytic scales by a reverse translation using the calculated meanic proximity ($P_{MnProx.}$) onto equations (14) and (15).

$$f: \begin{bmatrix} X = P_{MnProx} & \overset{\delta = 0}{\twoheadrightarrow} & P = 1 \\ \downarrow & \downarrow & \downarrow \\ R = r_1 & r_{n+1} & r_{2n+1} \end{bmatrix}$$

$$KC1_{Prox.}(X, P) = \frac{r_{n+1}(P_{MnProx} + 1)}{(r_1 \times P_{MnProx}) + (r_{2n+1} \times 1)}$$

Let $r = 1, 2, 3, \dots, 2n + 1$, and n is the sample size. That is in this case, the identity of the optical scale equals to the scale. Then we have:

$$KC1_{Prox.}(X, P) = \frac{n + 1\ (P_{MnProx} + 1)}{(1 \times P_{MnProx}) + (2n + 1 \times 1)} = \frac{n + 1\ (P_{MnProx} + 1)}{P_{MnProx} + (2n + 1)} \quad (14)$$

The alternate reflection and estimate of the above argument become

$$f: \begin{bmatrix} X = 1 & \overset{\delta = 0}{\twoheadrightarrow} & P = P_{MnProx} \\ \downarrow & \downarrow & \downarrow \\ R = r_1 & r_{n+1} & r_{2n+1} \end{bmatrix}$$

$$KC2_{Prox.}(X, P) = \frac{r_{n+1}(1 + P_{MnProx})}{(r_1 \times 1) + (r_{2n+1} \times P_{MnProx})}$$

Let $r = 1, 2, 3, \dots, 2n + 1$, and n is the sample size. That is in this case, the identity of the optical scale equals to the scale. Then we have:

$$KC2_{Prox.}(X, P) = \frac{n + 1\ (1 + P_{MnProx})}{(1 \times 1) + (2n + 1)P_{MnProx}} = \frac{n + 1\ (P_{MnProx} + 1)}{P_{MnProx}(2n + 1) + 1} \quad (15)$$

However, some parameters or properties, such as alternativeness, outlier biasedness, efficiency, and relative efficiency of the estimators were computed using the following statistics.

i. Alternativeness

Pearson correlation was used to correlate between the averages of the estimates through the range of variable standard deviations ($\sigma = 1, 2, 3, \dots, 15$), for each treated sample size ($n = 10, 50, 100, 200, \& 500$). A strong correlation between the proposed estimator and the gold standard represents suitability as an alternative method. Find the supplementary files (S1-S6).

ii. Robustness to contamination (Outlier biasedness)

The biasedness of the estimators under contaminations from a normal distribution was evaluated from the equation (16). Find the supplementary files (S7).

$$\text{Bias} = \text{E}(\varphi) - \text{E}(\hat{\varphi}) \quad (16)$$

Statisitcal mirroring

$$E(\varphi), (\hat{\varphi}) = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2}{\frac{1}{n}\sum_{i=1}^{n}(x)^2} = \frac{s^2}{x^2}$$

φ = A standardized estimate expectation of the estimator before contamination.

$\hat{\varphi}$ = A standardized estimate expectation of the estimator after contamination.

Note: x is the 1,000 estimates of the estimator.

iii. Relative absolute biasedness (RAB) of contaminations

Relative absolute biasedness of the estimator under contaminations is an evaluation that checks the equality of biasedness under positive and negative contaminations. It is expressed by the equation (17). Find the supplementary files (S7).

$$RAB\ of\ C = \frac{Bias\ of\ C^{+}}{Bias\ of\ C^{-}} \tag{17}$$

$Bias\ of\ C^{+}$ = Biasness of the estimator under positive contaminations.

$Bias\ of\ C^{-}$ = Biasness of the estimator under negative contaminations.

If $RAB = 1$, then the estimator has the same biasedness under positive (C^{+}) and negative (C^{-}) contaminations

If $RAB < 1$, then the estimator is less biased under positive contaminations

If $RAB > 1$, then the estimator is less biased under negative contaminations

iv. Normality of the distribution of the estimates of the estimators

Grapad Prism was used to calculate the P-value of the normality of the distribution of the estimates of the estimators. Find the supplementary files (S7).

v. Efficiency

The efficiency and relative efficiency of the estimators from a normal distribution was evaluated from equations (18) and (19) respectively. Find the supplementary (S1-S6).

$$Ef = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2}{\frac{1}{n}\sum_{i=1}^{n}(x)} = \frac{s^2}{x}$$

The standardized efficiency (standardized variance) now becomes:

$$Ef = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2}{\frac{1}{n}\sum_{i=1}^{n}(x)^2} = \frac{s^2}{x^2} \tag{18}$$

vi. Relative efficiency

$$REf = \frac{Ef(g)}{Ef(p)} \tag{19}$$

$Ef(g)$ = Efficiency of the gold (reference) standard estimator.

$Ef(p)$ = Efficiency of the proposed estimator.

If $REf = 1$, then both the estimators have same efficiency

If $REf < 1$, then the proposed estimator is less efficient

If $REf > 1$, then the proposed estimator is more efficient

5. Results and Discussion

5.1 Alternativeness of the estimators

Pearson correlation coefficient was used to check the alternativeness of the proposed estimators, the Kabirian coefficients of meanic proximity (KC1-MnProx., and KC2-MnProx.), the meanic proximity (MnProx.), and the meanic deviation (MnDev.) with the gold standard estimators of dispersion around the mean that functionalized with all the observations of the dataset from a normal distribution. The correlation coefficient between the estimates of two comparing methods (i.e, the proposed and reference estimates) for the order of parametrized standard deviations ($\sigma = 1, 2, 3, \dots, 15$) express the alternativeness of the estimator. All the mean estimates of the estimators are presented in Appendix A1.

The results in Table 2 shows that the proposed estimators (KC1-MnProx., KC2-MnProx., MnProx., MnDev.) are very strongly correlated and associated ($/R/ = 0.7674$ to 0.9994) with the standard deviation, coefficient of variation, and variance. The KC1-MnProx., and MnProx., were positively correlated with the standard deviation, coefficient of variation, and variance; while the KC2-MnProx., and MnDev. were negatively correlated. The statistical meanic mirroring is therefore a suitable alternative estimator of dispersion.

Table 2: Correlation coefficients between the estimators. Distribution [N ($\mu = 10;\ \sigma = 1, 2, 3, \dots, 15; n = 10, 50, 100, 200, 500;\ \&\ 1000\ random\ variables$)].

	Sample size/Correlation coefficient (R)						
Estimators	**10**	**50**	**100**	**200**	**500**	**/R/-Min.**	**/R/-Max.**
Proposed estimators' correlations with standard deviation (StDev.)							
KC1-MnProx. & StDev.	-0.9989	-0.9994	-0.9992	-0.9990	-0.9989	0.9989	0.9994
KC2-MnProx. & StDev.	0.9817	0.9946	0.9970	0.9975	0.9978	0.9817	0.9978
MnProx. & StDev.	-0.9992	-0.9994	-0.9991	-0.9990	-0.9989	0.9989	0.9994
MnDev. & StDev.	0.9947	0.9994	0.9991	0.9990	0.9989	0.9947	0.9994
Proposed estimators' correlations with a coefficient of variation (CV)							
KC1-MnProx. & CV.	-0.7675	-0.9985	-0.9987	-0.9988	-0.9988	0.7675	0.9988
KC2-MnProx. & CV.	0.8016	0.9968	0.9978	0.9980	0.9981	0.8016	0.9981
MnProx. & CV.	-0.7674	-0.9984	-0.9987	-0.9987	-0.9988	0.7674	0.9988
MnDev. & CV.	0.7778	0.9984	0.9987	0.9987	0.9988	0.7778	0.9988
Proposed estimators' correlations with variance (Var.)							
KC1-MnProx. & Var.	-0.9815	-0.9643	-0.9622	-0.9613	-0.9607	0.9607	0.9815
KC2-MnProx. & Var.	0.9782	0.9905	0.9874	0.9863	0.9856	0.9782	0.9905
MnProx. & Var.	-0.9798	-0.9640	-0.9621	-0.9612	-0.9606	0.9606	0.9798
MnDev. & Var.	0.9469	0.9640	0.9621	0.9612	0.9606	0.9469	0.9640

/R/-Min. = Absolute minimum correlation; /R/-Max. = Absolute maximum correlation

5.2 Outlier biasness and normality of the estimators' estimates under contaminations

Outlier biasedness of the Kabirian bi-coefficients of meanic proximity (KC1-MnProx., and KC2-MnProx.), the meanic proximity (MnProx.), and the meanic deviation (MnDev.) was compared with the standard deviation and coefficient of variation. A total of 1,000 artificial datasets from a normal distribution with $\mu = 10, \sigma = 2$ was contaminated with a varying magnitude of contaminants, and biasedness of the estimators due to the contaminants was analyzed. To simplify the graphical presentation of the results, the absolute outlier biasedness of the estimates was log-transformed. Therefore, the higher values of the log-transformed result represent low outlier biasedness and the vise-versa.

The Figure 1 and 2 presented how sensitive are the estimators to the contaminations (outliers). The results show that the statistical meanic mirroring (composed of KC1-MnProx., KC2-MnProx., MnProx., MnDev.) is less biased (less sensitive and more resistant to contaminations) than the standard deviation and coefficient of variation, at lower and extreme contaminations, from the top and the bottom points of the ordered random numbers. At the lower magnitude of contaminations, the negative contaminations lead to more bias than the positive contaminations for KC1-MnProx., KC2-MnProx., and MnProx. estimators; while MnDev., StDev., and CV are the opposite case. At the extreme magnitude of contaminations, both positive and negative contaminations are relatively the same outlier biasedness. In terms of the outlier biasedness between the possible Kabirian bi-coefficients, KC1-MnProx. is superior (more resistant) with negative contaminations than KC2-MnProx. with positive contaminations. Statistical mirroring is more sensitive to contamination as the mean of the distribution tends close to zero than away from zero mean, but the sensitivity is very less compared to the coefficient of variation.

Table B1 of Appendix B presented the normality distribution of the estimates of the estimators under contaminations (outliers). The results show that the estimates of the statistical meanic mirroring (KC1-MnProx., KC2-MnProx., MnProx., MnDev.) have passed the normality test under low and extreme outliers, except in the case of the mean tending close to zero due to negative outliers. In the case of StDev., the normality of the estimates has failed in all the examined cases of contaminations except in only one case. While for the case of coefficient of variation, the normality has given an imprecise result because it failed with low contaminations and passed with higher contaminations.

doi:10.20944/preprints201911.0268.v2

Statisitcal mirroring

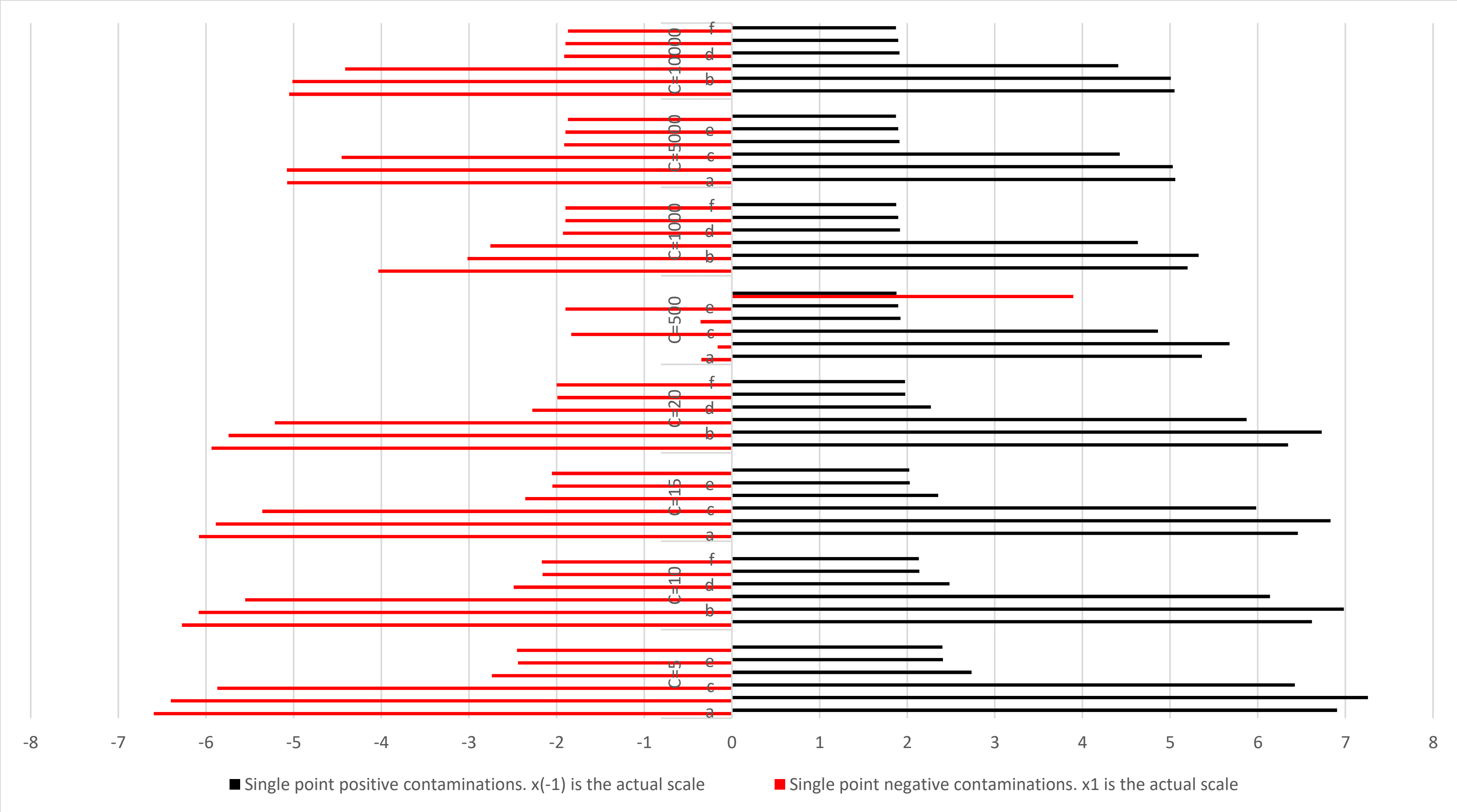

Figure 1: Log_{10} biasness of the estimators under single point contaminations from a normal distribution.

Keys: Letters **a-f** are the estimators (**a** = KC1-MnProx.; **b** = KC2-MnProx.; **c** = MnProx.; d = MnDev.; **e** = StDev.; **f** = CV); **C** = Contamination.

Statisitcal mirroring

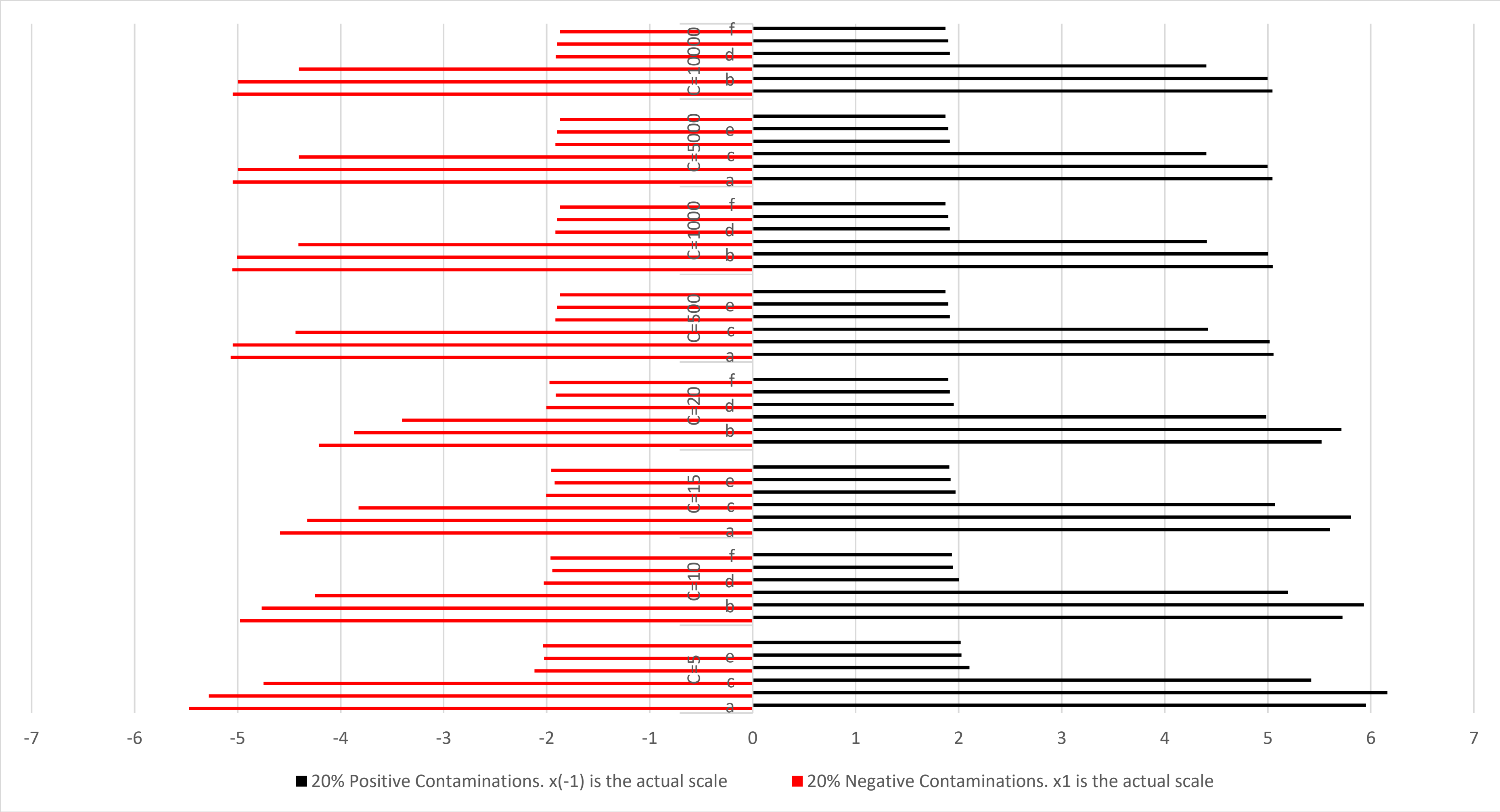

Figure 2: Log_{10} biasness of the estimators under 20% contaminations from a normal distribution.

Keys: Letters **a-f** are the estimators (**a** = KC1-MnProx.; **b** = KC2-MnProx.; **c** = MnProx.; d = MnDev.; **e** = StDev.; **f** = CV); **C** = Contamination.

5.3 Robustness (invariance) to scale

The scale invariance (scale robustness) of statistical mirroring has already been shown by its properties stated in this paper. In this case, attention is focused to compare the scale robustness of coefficient of variation and statistical meanic mirroring on two cases: a) a case of positive and negative scaling, b) a case of zero mean.

From the stated properties in this paper, a statistical mirroring is invariant to positive and negative scaling of a subset from natural numbers. It is therefore independent of whether the mean is positive or negative and all results are the same in all respect to scaling.

$$KC_{Prox.}(x,p) = KC_{Prox.}(ax,ap) = KC_{Prox.}(-ax,-ap)$$

$$P_{Prox.}(x,p) = P_{Prox.}(ax,ap) = P_{Prox.}(-ax,-ap)$$

$$P_{Dev.}(x,p) = P_{Dev.}(ax,ap) = P_{Dev.}(-ax,-ap)$$

where $x, p, a = \mathbb{R}; a \neq 0$. This means it satisfies the following network of relationships:

But the coefficient of variation (CV), it is either invariant to positive or negative scaling of a subset from natural numbers. Therefore, the invariance property of CV depends on whether the mean is positive or negative, all estimates are not the same in all respect of scaling.

$$\text{If } CV(x) \leq 0, \text{ then } CV(x) \neq CV(ax) = CV(-ax)$$

$$\text{If } CV(x) \geq 0, \text{ then } CV(x) = CV(ax) \neq CV(-ax)$$

where $x, p, a = \mathbb{R}; a \neq 0$.

In the case of a zero-mean from the set of integers or a constant zero value, the coefficient of variation (CV) functionally breaks down but statistical meanic mirroring does not. The functional breakdown of the CV is a result of zero denominators (i.e., zero mean) which is never found with the statistical mirroring except in the case of a uniform zero set of numbers. Even at this case, any small amount of optinalytic normalization can eliminate this scenario.

5.4 Robustness (invariance) to sample size

The sample size invariance (robustness to sample sizes) of statistical mirroring has already been shown by its properties stated in this paper. At this time, an example numerical problem (i.e., a measurement of relative diversity of a certain attribute) was provided (Table 3) to compare the impact of sample size on the estimators of dispersion. The results from Table 3 shows that, despite having an identical central tendency and score distribution, all the estimators (variance, standard deviation, coefficient of variation) are not robust to sample size except the statistical meanic mirroring (MnProx., and MnDev.).

In sociology and economics, the use of coefficient of variation to estimate demographic diversity index has been one of the potential problems for the comparison of groups with different sample sizes [11]. To make the groups comparable for their differing sample size, [13] created a corrected version of the coefficient of variation that resists sample size variation. Anthur & Kevin [11] reported that 29 out of 36 published articles from 1984 to 1999 on work group diversity have used uncorrected coefficient of variation as an index of diversity. Now, the simplest way to deal with this problem is the use of statistical meanic mirroring which is robust and unbiased to sample sizes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

Table 3: Impact of sample size on the estimators of dispersion

Group	n	Scores	Mean	StDev.	Var.	CV	MnProx.	MnDev.
A	3	5, 12, 22	13	8.54	73	0.66	0.86	0.14
B	6	5, 5 12, 12 22, 22	13	7.64	58.4	0.59	0.86	0.14
C	9	5, 5, 5 12, 12, 12 22, 22, 22	13	7.40	54.75	0.57	0.86	0.14
D	12	5, 5, 5, 5 12, 12, 12, 12 22, 22, 22, 22	13	7.29	53.09	0.56	0.86	0.14
E	24	5, 5, 5, 5 5, 5, 5, 5 12, 12, 12, 12, 12, 12, 12, 12 22, 22, 22, 22 22, 22, 22, 22	13	7.13	50.78	0.55	0.86	0.14

5.5 Efficiency and relative efficiency of the estimators

Efficiency and relative efficiency properties were used to evaluate the goodness of the statistical meanic mirroring (KC1-MnProx., KC2-MnProx., MnProx., MnDev.) and some gold standard estimators of dispersion around the mean. Total of 1000 artificial datasets from a normal distribution with $\mu = 10;\ \sigma = 1, 2, 3, \dots, 15; n = 10, 50, 100, 200, 500$ was used. The standardized variance of the estimates of the 1000 random numbers express the efficiency.

Tables 4 and 5 presented how efficient are the estimators. The results show that the estimators of statistical meanic mirroring (KC1-MnProx., KC2-MnProx., MnProx., MnDev.) are more or equally efficient as compared to standard deviation and coefficient of variation. Similar to the efficiency, the relative efficiency of all the estimators decreases with a higher spread of the normal distribution and leaves most of the estimators of statistical meanic mirroring the most superior efficiency. Fortunately, the superior efficiency gets declined or lost as the spread of the normal distribution gets wider. The loss in superior efficiency is due to the inconsistent and increased variance of the mean as the spread gets larger (Figure 3 and Appendix C1). Thus, this loss of efficiency is not an estimators' weakness per see, but it is a simulator's weakness. That is why only the two relative estimators of dispersion are affected because they are very relative to the mean of the distribution.

Table 4: Efficiency of the estimators from a normal distribution.

Distribution [N ($\mu = 10$; $\sigma = 1, 2, 3, \ldots, 15$; $n = 10, 50, 100, 200, 500$; & 1000 *random variables*)] and efficiency of the estimators																
n	Estimator	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$	$\sigma = 11$	$\sigma = 12$	$\sigma = 13$	$\sigma = 14$	$\sigma = 15$
10	**KC1-MnProx.**	0.0000	0.0000	0.0001	0.0002	0.0003	0.0005	0.0009	0.0012	0.0021	0.0037	0.0065	0.0107	0.0155	0.0180	0.0245
	KC2-MnProx.	0.0000	0.0000	0.0001	0.0002	0.0004	0.0008	0.0035	0.0026	0.4946	0.0981	0.5534	0.1714	2.1697	4.6842	1.0055
	MnProx.	0.0000	0.0002	0.0004	0.0008	0.0016	0.0027	0.0052	0.0072	0.0125	0.0227	0.0412	0.0687	0.1062	0.1312	0.1843
	MnDev.	0.0563	0.0681	0.0627	0.0666	0.0822	0.0891	0.1247	0.1275	0.1759	0.2207	0.2931	0.2793	0.4340	0.5566	0.5589
	StDev.	0.0551	0.0681	0.0579	0.0551	0.0567	0.0579	0.0668	0.0604	0.0580	0.0569	0.0545	0.0620	0.0602	0.0525	0.0546
	CV.	0.0560	0.0781	0.0670	0.0722	0.0924	0.1239	0.2089	0.1715	0.2811	0.7724	82.6822	638.4426	161.8866	19.0467	21.7340
50	**KC1-MnProx.**	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0002	0.0004	0.0005	0.0007	0.0009	0.0011	0.0015	0.0023
	KC2-MnProx.	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0004	0.0007	0.0009	0.0016	0.0020	0.0029	0.0047	0.0116
	MnProx.	0.0000	0.0000	0.0001	0.0002	0.0003	0.0005	0.0008	0.0012	0.0021	0.0026	0.0039	0.0052	0.0071	0.0098	0.0151
	MnDev.	0.0114	0.0109	0.0120	0.0126	0.0147	0.0155	0.0173	0.0195	0.0253	0.0239	0.0292	0.0307	0.0353	0.0384	0.0502
	StDev.	0.0122	0.0110	0.0110	0.0107	0.0122	0.0107	0.0106	0.0099	0.0120	0.0109	0.0111	0.0102	0.0115	0.0112	0.0113
	CV.	0.0124	0.0113	0.0132	0.0141	0.0190	0.0197	0.0216	0.0242	0.0345	0.0321	0.0438	0.0440	0.0525	0.0611	0.0859
100	**KC1-MnProx.**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0003	0.0003	0.0004	0.0005	0.0007	0.0009
	KC2-MnProx.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0003	0.0005	0.0007	0.0010	0.0011	0.0018	0.0024
	MnProx.	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003	0.0004	0.0006	0.0008	0.0014	0.0019	0.0027	0.0031	0.0045	0.0058
	MnDev.	0.0051	0.0053	0.0053	0.0061	0.0071	0.0082	0.0083	0.0089	0.0098	0.0130	0.0142	0.0158	0.0153	0.0184	0.0197
	StDev.	0.0053	0.0056	0.0052	0.0057	0.0057	0.0055	0.0055	0.0053	0.0051	0.0056	0.0060	0.0056	0.0058	0.0051	0.0054
	CV.	0.0054	0.0060	0.0061	0.0070	0.0087	0.0098	0.0104	0.0116	0.0131	0.0179	0.0199	0.0227	0.0222	0.0275	0.0304
200	**KC1-MnProx.**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0002	0.0003	0.0004	0.0004
	KC2-MnProx.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0003	0.0004	0.0006	0.0009	0.0011
	MnProx.	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0003	0.0004	0.0007	0.0008	0.0012	0.0017	0.0024	0.0029
	MnDev.	0.0026	0.0027	0.0028	0.0030	0.0035	0.0039	0.0041	0.0050	0.0052	0.0063	0.0062	0.0075	0.0083	0.0097	0.0099
	StDev.	0.0028	0.0027	0.0026	0.0028	0.0028	0.0028	0.0026	0.0028	0.0026	0.0028	0.0028	0.0025	0.0027	0.0026	0.0029
	CV.	0.0028	0.0030	0.0030	0.0036	0.0042	0.0047	0.0053	0.0064	0.0069	0.0084	0.0086	0.0104	0.0121	0.0144	0.0152

Table 4: Efficiency of the estimators from a normal distribution (Continuation)

		Distribution [N ($\mu = 10$; $\sigma = 1, 2, 3, \ldots, 15$; $n = 10, 50, 100, 200, 500$; & 1000 $random\ variables$)] and efficiency of the estimators														
n	Estimator	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$	$\sigma = 11$	$\sigma = 12$	$\sigma = 13$	$\sigma = 14$	$\sigma = 15$
500	**KC1-MnProx.**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0002
	KC2-MnProx.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.0004
	MnProx.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0009	0.0011
	MnDev.	0.0010	0.0011	0.0011	0.0012	0.0014	0.0014	0.0016	0.0019	0.0022	0.0025	0.0027	0.0030	0.0031	0.0035	0.0036
	StDev.	0.0010	0.0011	0.0011	0.0010	0.0011	0.0010	0.0011	0.0012	0.0012	0.0012	0.0011	0.0011	0.0011	0.0011	0.0011
	CV.	0.0011	0.0012	0.0013	0.0014	0.0017	0.0017	0.0020	0.0025	0.0028	0.0034	0.0037	0.0043	0.0044	0.0053	0.0055

 doi:10.20944/preprints201911.0268.v2

Table 5: Relative efficiency between statistical meanic mirroring and standard deviation from a normal distribution.

		Distribution [N ($\mu = 10$; $\sigma = 1,2,3,\ldots,15$; $n = 10,50,100,200,500$; & 1000 $random\ variables$)] and relative efficiency of the estimators														
n	Estimator	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$	$\sigma = 11$	$\sigma = 12$	$\sigma = 13$	$\sigma = 14$	$\sigma = 15$
10	**KC1-MnProx.**	7278.7588	1836.2836	742.9438	364.7208	198.0019	121.2064	73.1438	49.6345	28.0669	15.4146	8.3683	5.8060	3.8899	2.9216	2.2259
	KC2-MnProx.	6933.5124	1619.8241	635.4789	291.9936	147.3529	72.1822	19.2126	22.7889	0.1172	0.5800	0.0985	0.3616	0.0278	0.0112	0.0543
	MnProx.	1469.3113	362.3533	142.9232	68.4047	36.2814	21.6178	12.7854	8.4177	4.6230	2.5067	1.3240	0.9017	0.5672	0.4001	0.2963
	MnDev.	0.9785	0.9999	0.9233	0.8271	0.6897	0.6501	0.5355	0.4737	0.3295	0.2579	0.1859	0.2219	0.1388	0.0943	0.0977
50	**KC1-MnProx.**	5836.7492	1364.4076	543.8645	289.0332	179.1677	101.3328	66.8112	41.8125	30.5675	23.8380	16.3942	11.8125	10.0341	7.4089	5.0062
	KC2-MnProx.	5527.6384	1220.6153	460.3064	231.2518	133.9815	71.3042	44.0984	25.6829	16.7924	12.6473	7.0139	5.0177	3.9035	2.4008	0.9724
	MnProx.	1364.7611	310.2375	120.1767	62.1038	37.4004	20.4955	13.1298	7.9543	5.6326	4.2547	2.8312	1.9669	1.6214	1.1440	0.7473
	MnDev.	1.0719	1.0122	0.9155	0.8537	0.8268	0.6896	0.6106	0.5073	0.4725	0.4543	0.3804	0.3314	0.3247	0.2914	0.2253
100	**KC1-MnProx.**	5341.0350	1362.9443	577.7301	302.3027	167.0702	97.2740	69.8862	48.6253	33.0642	22.3221	17.7469	12.4252	11.4454	7.0941	6.0409
	KC2-MnProx.	5051.0053	1219.2994	487.8246	241.1729	125.2877	69.3316	46.8290	30.4728	19.5692	12.1109	9.0858	5.7257	5.0424	2.7946	2.1883
	MnProx.	1272.9192	315.7408	130.0522	66.0866	35.4486	20.0299	13.9528	9.4133	6.1989	4.0490	3.1122	2.1045	1.8789	1.1200	0.9188
	MnDev.	1.0257	1.0491	0.9992	0.9315	0.8071	0.6752	0.6623	0.6007	0.5187	0.4349	0.4206	0.3529	0.3766	0.2759	0.2715
200	**KC1-MnProx.**	5485.5837	1311.1832	526.6923	296.2489	162.6389	103.4100	66.5001	45.1377	31.6617	22.5867	18.7799	11.9197	9.7689	6.8314	6.5631
	KC2-MnProx.	5184.2290	1170.7700	445.3379	236.2761	122.4145	73.6785	44.7544	28.3336	18.9187	12.6400	9.8822	5.8323	4.4307	2.8551	2.5697
	MnProx.	1320.0533	306.6083	119.6888	65.3414	34.8169	21.4698	13.3921	8.8077	5.9873	4.1355	3.3249	2.0375	1.6111	1.0903	1.0088
	MnDev.	1.0739	1.0294	0.9211	0.9293	0.7967	0.7316	0.6380	0.5669	0.5016	0.4418	0.4463	0.3400	0.3294	0.2649	0.2945
500	**KC1-MnProx.**	5273.6491	1342.1248	564.2494	280.1364	166.0733	104.3343	68.6461	49.5277	33.9314	25.5617	16.8281	12.9194	10.7544	7.7002	6.8125
	KC2-MnProx.	4985.4896	1198.7299	476.5929	223.7237	125.2603	74.3787	46.1024	31.3972	20.2836	14.3998	8.9096	6.3900	5.0702	3.3903	2.8147
	MnProx.	1276.4703	315.6704	128.9017	62.1362	35.7389	21.7791	13.8948	9.7084	6.4422	4.7006	2.9908	2.2199	1.7854	1.2328	1.0529
	MnDev.	1.0435	1.0602	1.0029	0.8842	0.8218	0.7430	0.6644	0.6293	0.5453	0.5053	0.4049	0.3701	0.3618	0.3022	0.3064

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

Statisitcal mirroring

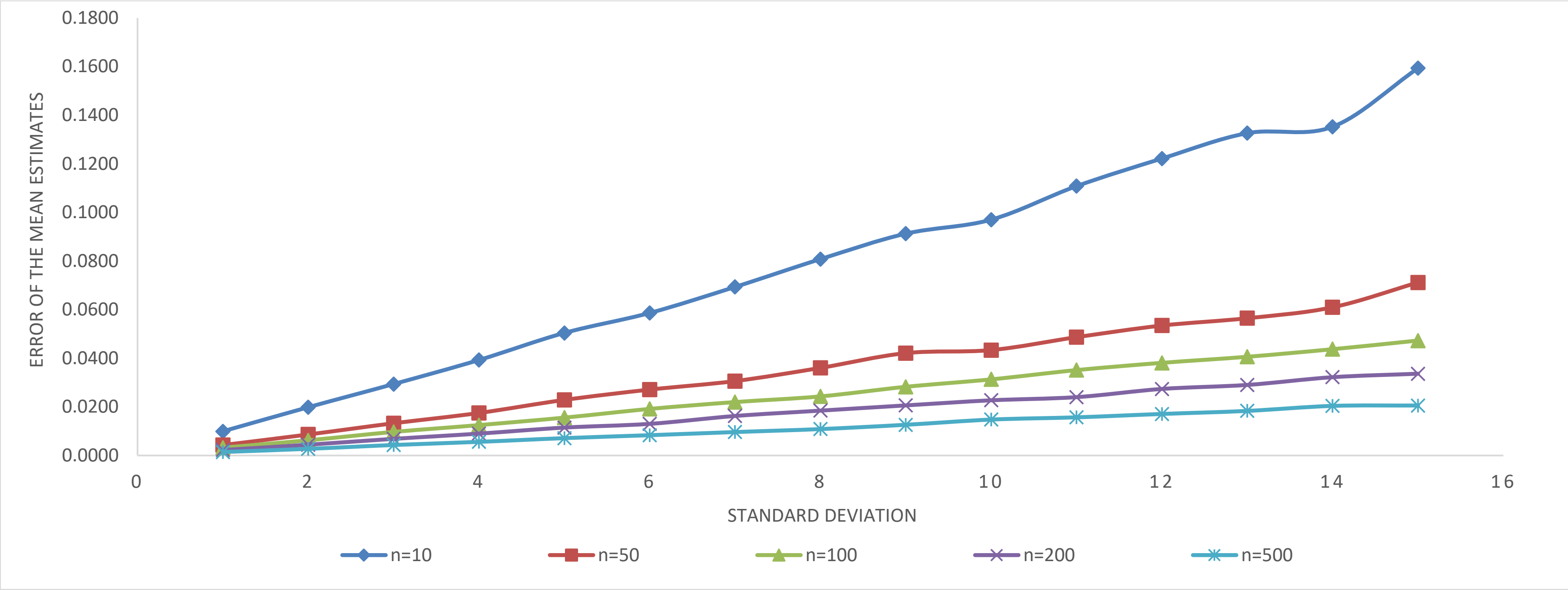

Figure 3: Monte Carlos simulation error of the mean from a normal distribution.

 doi:10.20944/preprints201911.0268.v2

6. Drawback and limitations of statistical meanic mirroring

The following are some of the identified drawbacks and limitations of statistical meanic mirroring:

i. Anonymity is not respected. It depends on the ordering of the list of the observations, otherwise inaccurate results.
ii. Under Gaussian distribution, it is more sensitive to contaminations as the mean tends close to zero than away from zero mean.
iii. It is inappropriate below a ratio level of scale.

7. Conclusion

Statistical mirroring (specifically the statistical meanic mirroring) is, under Gaussian distribution, a suitable alternative estimator of dispersion that is less biased (more resistant) to contaminations, robust to scale and sample size, and more efficient to a random distribution of variables as compared with some reference standard estimators. However, some of the limitations of statistical mirroring include: It relies on a) the ordering of the list of the observations, b) more sensitive to contaminations as the mean (of a normal distribution) tends close to zero, c) inappropriate below ratio level of scale.

7. Recommendation

In this paper, all the proposed estimators of statistical dispersion were not compared for their statistical goodness, evaluated for suitability and application. It is therefore recommended that other estimators of dispersion around defined location estimates or points (e.g, statistical medianic, maximalic, minimalic, modalic and rangic mirroring) should be explored for possible application and comparison.

The suitability and statistical goodness of statistical mirroring should also be checked for other distributions such as Poisson, uniform, binormial, chi-square, Bernoulli, patterned, discrete distributions, etc. However, the performance of the estimators with real datasets should be evaluated.

Supplementary material: The supplementary files attached are customized Excel sheets. Find the supplementary files (S1-S7).

Conflict of interest: The author declares no conflict of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

References

[1] Wada K. Outliers in official statistics. Japanese Journal of Statistics and Data Science (2020) 3:669-691. doi: 10.1007/s42081-020-00091-y

[2] Marazzi A. Improving the efficiency of robust estimators for the generalized linear model. Stats. (2021) 4:88-107. doi: 10.3390/stats4010008

[3] Félix MA, Barkoulas M. Pervasive robustness in biological systems. Nat Rev Genet. (2015) 16:483–96. doi: 10.1038/nrg3949

[6] Faber DS, Korn H. Applicability of the coefficient of variation method for analyzing synaptic plasticity. Biophys J. (1991) 60:1288–94. doi: 10.1016/S0006-3495(91)82162-2

[4] Gao Y, Vogt A, Forsyth CJ, Koide K. Comparison of splicing factor 3b inhibitors in human cells. ChemBioChem. (2013) 14:49–52. doi: 10.1002/cbic.201200558

[5] Lamare F, Le Maitre A, Dawood M, Schäfers KP, Fernandez P, Rimoldi OE, et al. Evaluation of respiratory and cardiac motion correction schemes in dual gated PET/CT cardiac imaging. Med Phys. (2014) 41:072504. doi: 10.1118/1.4881099

[7] Christodoulou C, Bugmann G. Coefficient of variation vs. mean interspike interval curves: what do they tell us about the brain? Neurocomputing. (2001) 38–40:1141–9. doi: 10.1016/S0925-2312(01)00480-5

[8] Castagliola P, Achouri A, Taleb H, Celano G, Psarakis S. Monitoring the coefficient of variation using a variable sampling interval control chart. Qual Reliabil Eng Int. (2013) 29:1135–49. doi: 10.1002/qre.1465

[9] Hulstijn JH, Van Gelderen A, Schoonen R. Automatization in second language acquisition: what does the coefficient of variation tell us? Appl Psycholinguist. (2009) 30:555–82. doi: 10.1017/S0142716409990014

[10] DerG,Deary IJ. Age and sex differences in reaction time in adulthood: results from the United Kingdom health and lifestyle survey. Psychol Aging. (2006) 21:62–73. doi: 10.1037/0882-7974.21.1.62

[11] Arthur GB, Kevin WM. On the use of coefficient of variation as a measure of diversity. Organizational Research Methods (2020) 3:285-297.

[12] Abdullahi KB. Optinalysis: A new method of data analysis and comparison. Preprints (2020) 2020080072. doi: 10.20944/preprints202008.0072.v2

[13] Smithson M. On relative dispersion: A new solution for some old problems. Quality and Quantity (1982) 16:261-271.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

Appendix A

Table A1: Average estimates of the estimators from a normal distribution.

		Distribution [N ($\mu = 10$; $\sigma = 1, 2, 3, \ldots, 15$; $n = 10, 50, 100, 200, 500$; & 1000 $random\ variables$)] and mean estimate of the estimators														
n	Estimator	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$	$\sigma = 11$	$\sigma = 12$	$\sigma = 13$	$\sigma = 14$	$\sigma = 15$
10	**KC1-MnProx.**	0.9886	0.9773	0.9661	0.9548	0.9446	0.9323	0.9221	0.9120	0.8987	0.8867	0.8728	0.8602	0.8437	0.8285	0.8194
	KC2-MnProx.	1.0117	1.0239	1.0366	1.0501	1.0631	1.0797	1.0959	1.1111	1.1599	1.1488	1.1397	1.2147	1.2106	1.2571	1.2637
	MnProx.	0.9748	0.9501	0.9256	0.9009	0.8788	0.8522	0.8301	0.8083	0.7797	0.7542	0.7247	0.6983	0.6636	0.6317	0.6129
	MnDev.	0.0252	0.0499	0.0744	0.0991	0.1212	0.1478	0.1699	0.1917	0.2183	0.2398	0.2613	0.2837	0.3004	0.3103	0.3171
	StDev.	0.9751	1.9579	2.9324	3.9440	4.8382	5.8629	6.7985	7.7158	8.7299	9.6457	10.6178	11.4505	12.7061	13.8388	14.5775
	CV.	9.7623	19.7320	29.6981	40.0694	49.7617	61.6433	72.2801	82.6116	96.5665	112.9791	99.2079	392.2186	227.7773	186.2883	174.9715
	Var.	1.0031	4.0943	9.0961	16.4118	24.7344	36.3635	49.3014	63.1244	80.6240	98.3304	118.8734	139.2322	171.1609	201.5549	224.0976
50	**KC1-MnProx.**	0.9866	0.9735	0.9606	0.9485	0.9365	0.9239	0.9129	0.9010	0.8897	0.8790	0.8682	0.8569	0.8480	0.8350	0.8256
	KC2-MnProx.	1.0137	1.0280	1.0428	1.0575	1.0730	1.0901	1.1060	1.1243	1.1430	1.1614	1.1819	1.2042	1.2238	1.2543	1.2829
	MnProx.	0.9727	0.9460	0.9197	0.8951	0.8706	0.8450	0.8226	0.7984	0.7754	0.7537	0.7318	0.7090	0.6908	0.6646	0.6456
	MnDev.	0.0273	0.0540	0.0803	0.1049	0.1294	0.1550	0.1774	0.2016	0.2246	0.2463	0.2682	0.2910	0.3092	0.3354	0.3544
	StDev.	0.9943	1.9956	3.0067	3.9840	4.9788	6.0080	6.9508	7.9561	8.9784	9.9595	10.9498	11.9766	12.9152	13.9694	14.9115
	CV.	9.9417	19.9874	30.1285	39.9149	49.9640	60.6505	70.2445	80.7646	91.4471	101.4962	112.2180	123.3242	132.8203	146.6056	157.3074
	Var.	1.0006	4.0261	9.1393	16.0424	25.0906	36.4818	48.8239	63.9247	81.5747	100.2667	121.2274	144.8971	168.7110	197.3266	224.8658
100	**KC1-MnProx.**	0.9863	0.9730	0.9601	0.9474	0.9351	0.9231	0.9114	0.9001	0.8888	0.8777	0.8667	0.8560	0.8467	0.8355	0.8253
	KC2-MnProx.	1.0141	1.0285	1.0434	1.0588	1.0747	1.0910	1.1080	1.1253	1.1435	1.1629	1.1830	1.2042	1.2234	1.2486	1.2731
	MnProx.	0.9724	0.9455	0.9194	0.8939	0.8689	0.8449	0.8211	0.7983	0.7756	0.7532	0.7312	0.7095	0.6908	0.6683	0.6478
	MnDev.	0.0276	0.0545	0.0806	0.1061	0.1311	0.1551	0.1789	0.2017	0.2244	0.2468	0.2688	0.2905	0.3092	0.3317	0.3522
	StDev.	1.0001	2.0007	2.9987	4.0026	4.9990	5.9974	6.9862	7.9671	9.0204	9.9991	11.0105	11.9766	12.9349	13.9695	14.9988
	CV.	10.0029	20.0224	30.0197	40.0490	50.1929	60.1621	70.2972	80.2825	90.4226	100.8287	111.3351	121.8727	131.2600	142.7067	153.6004
	Var.	1.0054	4.0251	9.0397	16.1112	25.1328	36.1684	49.0758	63.8143	81.7794	100.5468	121.9566	144.2392	168.2749	196.1369	226.1672

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

Table A1: Average estimates of the estimators from a normal distribution (Continuation).

		Distribution [N ($\mu = 10$; $\sigma = 1, 2, 3, \ldots, 15$; $n = 10, 50, 100, 200, 500$; & 1000 $random$ $variables$)] and mean estimate of the estimators														
n	Estimator	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$	$\sigma = 11$	$\sigma = 12$	$\sigma = 13$	$\sigma = 14$	$\sigma = 15$
200	**KC1-MnProx.**	0.9862	0.9727	0.9599	0.9470	0.9346	0.9224	0.9108	0.8993	0.8883	0.8774	0.8665	0.8556	0.8450	0.8356	0.8253
	KC2-MnProx.	1.0142	1.0288	1.0436	1.0593	1.0753	1.0919	1.1087	1.1263	1.1442	1.1629	1.1827	1.2039	1.2259	1.2467	1.2707
	MnProx.	0.9723	0.9452	0.9194	0.8934	0.8686	0.8442	0.8208	0.7976	0.7755	0.7537	0.7319	0.7100	0.6886	0.6698	0.6492
	MnDev.	0.0277	0.0548	0.0806	0.1066	0.1314	0.1558	0.1792	0.2024	0.2245	0.2463	0.2681	0.2900	0.3114	0.3302	0.3508
	StDev.	1.0020	2.0057	2.9928	3.9997	5.0031	6.0045	7.0004	8.0059	8.9898	10.0052	11.0076	12.0154	12.9935	14.0585	14.9750
	CV.	10.0169	20.0557	29.9239	40.1040	50.0933	60.1840	70.1189	80.2135	90.1589	100.1479	110.4628	120.9149	131.5974	141.1092	151.8977
	Var.	1.0067	4.0339	8.9801	16.0420	25.1004	36.1568	49.1355	64.2765	81.0275	100.3831	121.5033	144.7374	169.2913	198.1485	224.9058
500	**KC1-MnProx.**	0.9861	0.9727	0.9595	0.9468	0.9343	0.9222	0.9104	0.8987	0.8875	0.8768	0.8657	0.8553	0.8451	0.8347	0.8251
	KC2-MnProx.	1.0143	1.0289	1.0440	1.0595	1.0757	1.0922	1.1091	1.1271	1.1453	1.1637	1.1838	1.2040	1.2250	1.2476	1.2699
	MnProx.	0.9722	0.9452	0.9189	0.8934	0.8683	0.8441	0.8206	0.7971	0.7746	0.7531	0.7310	0.7101	0.6896	0.6688	0.6496
	MnDev.	0.0278	0.0548	0.0811	0.1066	0.1317	0.1559	0.1794	0.2029	0.2254	0.2469	0.2690	0.2899	0.3104	0.3312	0.3504
	StDev.	1.0015	2.0008	3.0027	4.0017	5.0060	6.0020	7.0101	8.0170	8.9976	10.0076	11.0141	12.0110	13.0062	14.0194	15.0136
	CV.	10.0162	20.0125	30.0281	39.9949	50.0914	60.0995	70.0439	80.3110	90.2666	100.1523	110.4899	120.5712	130.6445	141.1313	151.1491
	Var.	1.0040	4.0075	9.0260	16.0304	25.0886	36.0598	49.1942	64.3487	81.0529	100.2763	121.4424	144.4275	169.3521	196.7534	225.6584

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

Appendix B

Table B1: Bias and relative absolute bias due to contaminations, of the statistical meanic mirroring and some estimators. Distribution $[N\ (\mu = 10, \sigma = 2\ \&\ 1000\ random\ variables)]$.

			Single Point PC		Single Point NC			20% PC		20% NC		
C	**Estimators**	**BC**	**AC**	**Bias**	**AC**	**Bias**	**RAB**	**AC**	**Bias**	**AC**	**Bias**	**RAB**
±5	**KC1-MnProx.**	8.9774E-06	8.8530E-06	1.2436E-07 NP	9.2313E-06	-2.5392E-07 NP	0.4897	7.8643E-06	1.1130E-06 NP	1.2366E-05	-3.3888E-06 NP	0.3284
	KC2-MnProx.	1.0029E-05	9.9737E-06	5.5196E-08 NP	1.0425E-05	-3.9625E-07 NP	0.1393	9.3409E-06	6.8797E-07 NP	1.5266E-05	-5.2370E-06 NP	0.1314
	MnProx.	3.9473E-05	3.9096E-05	3.7697E-07 NP	4.0816E-05	-1.3433E-06 NP	0.2806	3.5701E-05	3.7722E-06 NP	5.7266E-05	-1.7793E-05 NP	0.2120
	MnDev.	1.2199E-02	1.0354E-02	1.8445E-03 NP	1.0373E-02	1.8260E-03 NP	1.0101	4.3515E-03	7.8470E-03 NP	4.5771E-03	7.6214E-03 NP	1.0296
	StDev.	1.2607E-02	8.7054E-03	3.9016E-03 NF	8.9746E-03	3.6324E-03 NF	1.0741	3.2095E-03	9.3974E-03 NF	3.1987E-03	9.4083E-03 NF	0.9988
	CV.	1.3449E-02	9.4828E-03	3.9662E-03 NF	9.9126E-03	3.5364E-03 NF	1.1216	3.8671E-03	9.5819E-03 NF	4.2440E-03	9.2050E-03 NP	1.0410
±10	**KC1-MnProx.**	8.9774E-06	8.7370E-06	2.4039E-07 NP	9.5069E-06	-5.2955E-07 NP	0.4539	7.0933E-06	1.8841E-06 NP	1.9438E-05	-1.0460E-05 NP	0.1801
	KC2-MnProx.	1.0029E-05	9.9246E-06	1.0424E-07 NP	1.0856E-05	-8.2716E-07 NP	0.1260	8.8667E-06	1.1621E-06 NP	2.7205E-05	-1.7176E-05 NP	0.0677
	MnProx.	3.9473E-05	3.8749E-05	7.2400E-07 NP	4.2276E-05	-2.8025E-06 NP	0.2583	3.3083E-05	6.3906E-06 NP	9.6239E-05	-5.6766E-05 NP	0.1126
	MnDev.	1.2199E-02	8.9111E-03	3.2874E-03 NP	8.9498E-03	3.2488E-03 NP	1.0119	2.3165E-03	9.8821E-03 NP	2.8161E-03	9.3825E-03 NP	1.0532
	StDev.	1.2607E-02	5.3244E-03	7.2825E-03 NF	5.6819E-03	6.9250E-03 NF	1.0516	1.2557E-03	1.1351E-02 NF	1.2486E-03	1.1358E-02 NF	0.9994
	CV.	1.3449E-02	6.0704E-03	7.3786E-03 NF	6.6696E-03	6.7794E-03 NF	1.0884	1.8122E-03	1.1637E-02 NP	2.5797E-03	1.0869E-02 NF	1.0706
±15	**KC1-MnProx.**	8.9774E-06	8.6284E-06	3.4891E-07 NP	9.8060E-06	-8.2866E-07 NP	0.4210	6.4906E-06	2.4868E-06 NP	3.4867E-05	-2.5889E-05 NP	0.0961
	KC2-MnProx.	1.0029E-05	9.8811E-06	1.4777E-07 NP	1.1325E-05	-1.2961E-06 NP	0.1140	8.4722E-06	1.5567E-06 NP	5.7455E-05	-4.7426E-05 NF	0.0328
	MnProx.	3.9473E-05	3.8429E-05	1.0441E-06 NP	4.3861E-05	-4.3880E-06 NP	0.2379	3.0981E-05	8.4920E-06 NP	1.8882E-04	-1.4935E-04 NP	0.0569
	MnDev.	1.2199E-02	7.7609E-03	4.4376E-03 NP	7.8210E-03	4.3776E-03 NP	1.0137	1.4815E-03	1.0717E-02 NP	2.2997E-03	9.8989E-03 NP	1.0827
	StDev.	1.2607E-02	3.2848E-03	9.3222E-03 NF	3.6282E-03	8.9788E-03 NF	1.0382	6.4780E-04	1.1959E-02 NF	6.4275E-04	1.1964E-02 NP	0.9996
	CV.	1.3449E-02	4.0154E-03	9.4336E-03 NF	4.6408E-03	8.8082E-03 NF	1.0710	1.1259E-03	1.2323E-02 NP	2.3894E-03	1.1060E-02 NP	1.1142
±20	**KC1-MnProx.**	8.9774E-06	8.5267E-06	4.5064E-07 NP	1.0131E-05	-1.1532E-06 NP	0.3908	5.9835E-06	2.9938E-06 NP	7.0518E-05	-6.1541E-05 NP	0.0486
	KC2-MnProx.	1.0029E-05	9.8425E-06	1.8638E-07 NP	1.1836E-05	-1.8068E-06 NP	0.1032	8.1060E-06	1.9229E-06 NP	1.4518E-04	-1.3515E-04 NF	0.0142
	MnProx.	3.9473E-05	3.8133E-05	1.3400E-06 NP	4.5585E-05	-6.1114E-06 NP	0.2193	2.9141E-05	1.0333E-05 NP	4.3343E-04	-3.9396E-04 NP	0.0262
	MnDev.	1.2199E-02	6.8290E-03	5.3696E-03 NP	6.9114E-03	5.2871E-03 NP	1.0156	1.0498E-03	1.1149E-02 NP	2.2601E-03	9.9385E-03 NP	1.1218
	StDev.	1.2607E-02	2.1281E-03	1.0479E-02 NF	2.4253E-03	1.0182E-02 NF	1.0292	3.9122E-04	1.2216E-02 NF	3.8748E-04	1.2219E-02 NF	0.9997

 doi:10.20944/preprints201911.0268.v2

Statisitcal mirroring

	CV.	1.3449E-02	2.8490E-03	1.0600E-02 [NF]	3.4539E-03	9.9951E-03 [NF]	1.0605	8.0594E-04	1.2643E-02 [NP]	2.7765E-03	1.0672E-02 [NF]	1.1846
±500	**KC1-MnProx.**	8.9774E-06	4.6280E-06	4.3494E-06 [NP]	4.4923E-01	-4.4922E-01 [NF]	0.0000	2.1501E-07	8.7623E-06 [NP]	4.2285E-07	8.5545E-06 [NP]	1.0243
	KC2-MnProx.	1.0029E-05	7.9312E-06	2.0976E-06 [NP]	6.8731E-01	-6.8730E-01 [NF]	0.0000	4.4659E-07	9.5823E-06 [NP]	1.0384E-06	8.9904E-06 [NP]	1.0658
	MnProx.	3.9473E-05	2.5698E-05	1.3775E-05 [NP]	1.4697E-02	-1.4658E-02 [NF]	0.0009	1.3367E-06	3.8136E-05 [NP]	2.9111E-06	3.6562E-05 [NP]	1.0431
	MnDev.	1.2199E-02	2.5788E-04	1.1941E-02 [NP]	4.5064E-01	-4.3844E-01 [NF]	0.0272	6.5440E-06	1.2192E-02 [NP]	8.6040E-06	1.2190E-02 [NP]	1.0002
	StDev.	1.2607E-02	3.5712E-06	1.2603E-02 [NF]	5.0803E-06	1.2602E-02 [NF]	1.0001	7.4553E-07	1.2606E-02 [NF]	7.3004E-07	1.2606E-02 [NF]	1.0000
	CV.	1.3449E-02	2.1436E-04	1.3235E-02 [NP]	7.8597E+03	-7.8597E+03 [NF]	0.0000	7.6464E-06	1.3441E-02 [NP]	1.1252E-05	1.3438E-02 [NP]	1.0003
±1000	**KC1-MnProx.**	8.9774E-06	2.6708E-06	6.3065E-06 [NP]	1.0144E-04	-9.2464E-05 [NP]	0.0682	6.1934E-08	8.9154E-06 [NP]	8.7817E-08	8.8895E-06 [NP]	1.0029
	KC2-MnProx.	1.0029E-05	5.2912E-06	4.7376E-06 [NP]	9.7063E-04	-9.6060E-04 [NF]	0.0049	1.3259E-07	9.8963E-06 [NP]	2.0416E-07	9.8247E-06 [NP]	1.0073
	MnProx.	3.9473E-05	1.6140E-05	2.3333E-05 [NP]	1.7941E-03	-1.7546E-03 [NP]	0.0133	3.9210E-07	3.9081E-05 [NP]	5.8451E-07	3.8889E-05 [NP]	1.0049
	MnDev.	1.2199E-02	9.2801E-05	1.2106E-02 [NP]	3.8782E-04	1.1811E-02 [NP]	1.0250	1.7408E-06	1.2197E-02 [NP]	2.0192E-06	1.2197E-02 [NP]	1.0000
	StDev.	1.2607E-02	8.8701E-07	1.2606E-02 [NF]	1.2696E-06	1.2606E-02 [NF]	1.0000	1.8693E-07	1.2607E-02 [NF]	1.8297E-07	1.2607E-02 [NF]	1.0000
	CV.	1.3449E-02	9.4780E-05	1.3354E-02 [NP]	8.5346E-04	1.2596E-02 [NF]	1.0602	2.0818E-06	1.3447E-02 [NP]	2.5439E-06	1.3446E-02 [NP]	1.0000
±5000	**KC1-MnProx.**	8.9774E-06	2.7662E-07	8.7007E-06 [NP]	5.4873E-07	8.4286E-06 [NP]	1.0323	2.7907E-09	8.9746E-06 [NP]	3.0508E-09	8.9743E-06 [NP]	1.0000
	KC2-MnProx.	1.0029E-05	6.7910E-07	9.3497E-06 [NP]	1.6602E-06	8.3686E-06 [NP]	1.1172	6.1345E-09	1.0023E-05 [NP]	6.8169E-09	1.0022E-05 [NP]	1.0001
	MnProx.	3.9473E-05	1.9040E-06	3.7569E-05 [NP]	4.3115E-06	3.5162E-05 [NP]	1.0685	1.7952E-08	3.9455E-05 [NP]	1.9821E-08	3.9453E-05 [NP]	1.0000
	MnDev.	1.2199E-02	5.6323E-06	1.2193E-02 [NP]	7.4684E-06	1.2191E-02 [NP]	1.0002	7.3362E-08	1.2198E-02 [NP]	7.7046E-08	1.2198E-02 [NP]	1.0000
	StDev.	1.2607E-02	3.5293E-08	1.2607E-02 [NF]	5.0756E-08	1.2607E-02 [NF]	1.0000	7.4947E-09	1.2607E-02 [NF]	7.3336E-09	1.2607E-02 [NF]	1.0000
	CV.	1.3449E-02	7.0320E-06	1.3442E-02 [NP]	1.0486E-05	1.3439E-02 [NP]	1.0003	8.9467E-08	1.3449E-02 [NP]	9.4296E-08	1.3449E-02 [NP]	1.0000
±10000	**KC1-MnProx.**	8.9774E-06	8.0041E-08	8.8973E-06 [NP]	1.1384E-07	8.8635E-06 [NP]	1.0038	7.0840E-10	8.9766E-06 [NP]	7.4971E-10	8.9766E-06 [NP]	1.0000
	KC2-MnProx.	1.0029E-05	2.0428E-07	9.8246E-06 [NP]	3.2204E-07	9.7068E-06 [NP]	1.0121	1.5626E-09	1.0027E-05 [NP]	1.6673E-09	1.0027E-05 [NP]	1.0000
	MnProx.	3.9473E-05	5.6442E-07	3.8909E-05 [NP]	8.5667E-07	3.8616E-05 [NP]	1.0076	4.5666E-09	3.9469E-05 [NP]	4.8569E-09	3.9468E-05 [NP]	1.0000
	MnDev.	1.2199E-02	1.5012E-06	1.2197E-02 [NP]	1.7465E-06	1.2197E-02 [NP]	1.0000	1.8464E-08	1.2199E-02 [NP]	1.9152E-08	1.2199E-02 [NP]	1.0000
	StDev.	1.2607E-02	8.8175E-09	1.2607E-02 [NF]	1.2688E-08	1.2607E-02 [NF]	1.0000	1.8742E-09	1.2607E-02 [NF]	1.8339E-09	1.2607E-02 [NF]	1.0000
	CV.	1.3449E-02	1.9290E-06	1.3447E-02 [NP]	2.3530E-06	1.3447E-02 [NP]	1.0000	2.2573E-08	1.3449E-02 [NP]	2.3357E-08	1.3449E-02 [NP]	1.0000

Key: C = Contamination; PC = Positive Contamination; NC = Negative Contamination; BC = Before Contamination; AC = After Contamination; RAB = Relative Absolute Bias; [NP]= Normality Passed; [NF]= Normality Failed.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2021 doi:10.20944/preprints201911.0268.v2

Appendix C

Table C1: Monte Carlos simulation error of the mean from a normal distribution.

Distribution [N ($\mu = 10$; $\sigma = 1, 2, 3, \dots, 15$; $n = 10, 50, 100, 200, 500$; & 1000 *random variables*)] and error of the mean of simulation															
n	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 6$	$\sigma = 7$	$\sigma = 8$	$\sigma = 9$	$\sigma = 10$	$\sigma = 11$	$\sigma = 12$	$\sigma = 13$	$\sigma = 14$	$\sigma = 15$
10	0.0100	0.0199	0.0294	0.0393	0.0504	0.0586	0.0693	0.0808	0.0913	0.0970	0.1108	0.1222	0.1326	0.1352	0.1593
50	0.0043	0.0086	0.0133	0.0175	0.0229	0.0271	0.0306	0.0360	0.0421	0.0434	0.0487	0.0535	0.0565	0.0610	0.0712
100	0.0032	0.0063	0.0097	0.0125	0.0156	0.0192	0.0220	0.0243	0.0283	0.0313	0.0351	0.0381	0.0406	0.0437	0.0473
200	0.0023	0.0044	0.0069	0.0090	0.0115	0.0130	0.0163	0.0185	0.0206	0.0227	0.0240	0.0274	0.0290	0.0322	0.0336
500	0.0014	0.0027	0.0043	0.0056	0.0071	0.0084	0.0097	0.0109	0.0126	0.0148	0.0157	0.0171	0.0184	0.0204	0.0206