



HAL
open science

By Means of the Means: Arithmetic, Harmonic, Geometric ...

Luciano da Fontoura Costa

► **To cite this version:**

Luciano da Fontoura Costa. By Means of the Means: Arithmetic, Harmonic, Geometric 2023.
hal-04152919

HAL Id: hal-04152919

<https://hal.science/hal-04152919v1>

Preprint submitted on 5 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

By Means of the Means: Arithmetic, Harmonic, Geometric, ...

Luciano da F. Costa
luciano@ifsc.usp.br

São Carlos Institute of Physics – DFCM/USP

17th Jun. 2023

Abstract

Though having no formal mathematical definition other than corresponding to a functional mapping a set of values into a respective scalar having the same physical unit, the concept of means play a fundamental role in science and technology for its ability to summarize data as well as an indication of a respective central tendency. In addition, the arithmetic mean, including its weighted version, coincide with the physical concept of center of mass. After revising the concepts of center of mass, fiunctionals, and statistical expectance, the present work aims at presenting some of the main types of means — namely the arithmetic, weighted, truncated, geometric, harmonic, Lehmer, contraharmonic and root mean square, as well as some of their main respective characteristics. The possibility to combine more than one type of mean as a means to characterize datasets is also briefly outlines and illustrated.

1 Introduction

We live in a world that increasingly relies on generating, analyzing and modeling *data* (e.g. [1]). In particular, given that collected data sets can have large sizes, it becomes interesting to conceive means for *summarizing* these sets in terms of a substantially smaller number of respective numeric properties.

Arguably, the *mean* of a set of numeric values (e.g. [2, 3, 4, 5, 6, 7, 8]) corresponds to the most natural and frequently adopted quantity employed to summarize the set from which it is calculated, as illustrated in Figure 1.

Such special importance and popularity of the mean can be traced back to antiquity, to the point that the arithmetic, geometric and harmonic means are often referred to as *Pythagorean means*, as they were studied as early as by the Pythagoreans under a mainly *geometric* point of view. The understanding of the mean as a characterization of the *centrality* of data emanates particularly from this type of geometric approaches, and especially from the directly related concept of *center of mass* (e.g. [9, 10, 11, 12]).

The importance of the concept of mean is also reflected in its several types which have been developed along millennia, which includes but are not limited to *arithmetic*, *geometric*, *harmonic*, and *root mean square*. Many relevant issues are therefore implied by the existence of several types of means, including in which sense they differ one another, which are their properties, and when it would

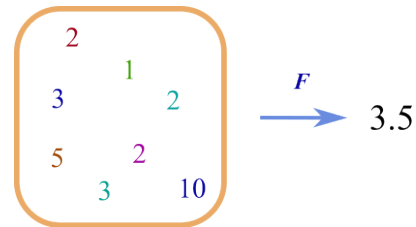


Figure 1: The arithmetic mean as a means to summarize a set of discrete integer values. Observe that the mean constitutes a *functional*, mapping a set of numbers into a single scalar value having the same physical unit.

be interesting to apply each of them.

One of the main objectives of the present work is to address some of the main types of means from the above perspectives. However, given the special role of the mean as a means for data characterization and summarization, we also pay special attention to how the considered types can be applied, and eventually combined, for that finality.

We start by discussing the meaning of the concept of mean and by presenting some of the main types of data that are typically available, and which can be respectively characterized in terms of means. Then, we provide a brief introduction to the concept of center of mass in terms of Newtonian mechanics, and then from the perspective of probability and statistics. The several considered types of mean — including arithmetic, weighted, truncated, geo-

metric, harmonic, Lehmer, contraharmonic, as well as the root mean square — are introduced, discussed and illustrated. The possibility to combine more than one type of means as a means of characterizing and representing data is then briefly discussed.

2 What is the Meaning of *Mean*?

Before proceeding further with our discussion about the mean, it is of particular interest to contemplate its meaning and possible definition. It may come as a surprise that, unlikely other mathematical concepts including *distances* (which need to satisfy some requisites), there is no formal definition of what a mean actually is, except that it necessarily corresponds to a *functional*, namely a mapping from the data into a single scalar (typically real) value having the same physical unit. For instance, a mean of distance values represented in meters will yield a scalar also in meters.

Informally speaking, the mean has been frequently associated to the ideas of *centrality* and *summarization/expectance*.

The connotation of *centrality* can be readily appreciated by considering its intrinsic relationship with the physical concept of *centroid*, or *center of mass*, of a given object, which can be discrete (a finite set of point masses) or continuous (an infinite set of connected point masses). As a matter of fact, the means of the coordinates of a set of points, each with the same mass, along each of the involved axes corresponds formally to the *center of the mass* of that object.

As discussed further in Section 4, the center of mass of a 3D (or 2D) object represents the point at which the respective object will undergo only linear acceleration, and no angular acceleration, while affected by a force passing through it. Alternatively, an object can be placed into unstable equilibrium when placed onto a wedge directed at its center of mass.

The mean therefore conveys the idea of corresponding to the ‘most central point’ of an object, taken from the specific physical-geometric perspective and, as such, it has some special properties as above indicated.

The connotation of the mean as *summarization* and *expectance* is typically related, as it will be further discussed in Section 5, to the areas of data analysis, statistics and probability. From this perspective, the mean of a data set would provide an effective summarized description of the respective data elements which would be, however, incomplete in the sense that several other properties of the data could not be taken into account by a single scalar measurement. For instance, a lot of apples can be effectively summarized in terms of the respective average weight, or

of some other respective property of particular interest (e.g. diameter or Brix sweetness).

At the same time, the arithmetic mean can also be probabilistic understood as being related to the *expectance* of the random variable from which the data elements were sampled. Observe, however, that expectance and the arithmetic mean, however, are not identical concepts. More strictly speaking, the arithmetic mean of a set of available samples provides an *estimation* of the expectance based of the respective continuous random variable.

So, we can summarize the above considerations as:

A *mean* of a set of numeric observations is henceforth understood as corresponding to a *functional* mapping that set into a scalar having the *same physical unit*, while *summarizing* the data and/or providing some information about its *centrality*.

Though often used indistinctly, the mean is also not the same as *average*, in the sense that the latter typically associated to the arithmetic mean, while the concept of mean is broader and encompasses several other types.

3 Types of Data

Given that we are particularly interested in using means for characterizing and representing data, it is interesting to consider some respective possible *types*. To being with, we observe that by *data* we will be referring to *collections* of respective *data elements*. In the case of numeric data, the values of the observed data elements is often called *scores* or *raw scores*.

Basically, we shall understand the two following overall types of *data*: (i) *unstructured* and (ii) *structured*. Unstructured datasets will henceforth be taken to correspond to the traditional concept of *sets*, being therefore characterized by the fact that the order of the data elements does not matter. On their hand, structured data involve not only the data elements themselves, but also their mutual *relationships* (including order), as is the case of discrete data structures including vectors, matrices, lists, trees, graphs, as well as continuous structures such as functions as well as scalar and vector fields.

Regarding the number of elements in a data, we can have *finite data* or *infinite data*, both of which will be considered in the present work. Continuous data are necessarily infinite, but discrete data sets can also be infinite.

Another possible division of types of data is as *categorical* and *numeric*. The former case includes data elements corresponding to generic symbols, including labels,

words, tags, etc. As implied in their own name, numeric data elements are numbers of several types, including natural, integer, rational, irrational, real, complex, etc. In the present work we will focus on numeric data, though categories can also be translated into numeric values by imposing some respective mapping.

In principle, the mean can be estimated from any data involving numeric entries. However, as the mean does not consider any of the relationships (including eventual order) between the data elements, any additional relational information originally available in the original data will not be taken into account. In such cases, it often becomes interesting to complement the mean in terms of additional functionals capable of yielding more comprehensive summarizations of the original relational data.

4 The Center of Mass

In Newtonian physics (e.g. [9, 10, 11, 12]), the concept of center of mass is intrinsically related to the idea of *angular acceleration* and *torque*. In the case of a point particle with mass m and a reference point P , as illustrated in Figure 2, the respective torque $\vec{\tau}$ can be expressed as:

$$\vec{\tau} = \vec{F} \times \vec{r} \quad (1)$$

where \times is the *cross product* between two vectors. It follows that:

$$|\vec{\tau}| = |\vec{F}| |\vec{r}| \sin(\alpha) = |\vec{F}| d \quad (2)$$

where the distance d is often called the *lever arm*, and α is the smaller angle between the vectors \vec{F} and \vec{r} .

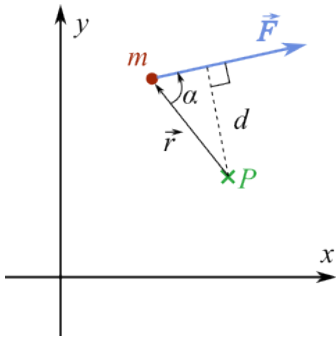


Figure 2: A point particle with mass m is under action of a force F as shown in the diagram. The *torque* of the force respectively to a reference point P , defining the relative position vector r to the particle, corresponds to the quantity $\vec{\tau} = \vec{F} \times \vec{r}$. It follows from properties of the the cross product that the the magnitude of τ can be calculated as the product between the magnitude of \vec{F} and the *lever arm* $d = |\vec{r}| \sin(\alpha)$.

Let us consider the physical arrangement of N point particles, all with the same mass m , mechanically vinculated linear along the x -axis, as illustrated in Figure 3.

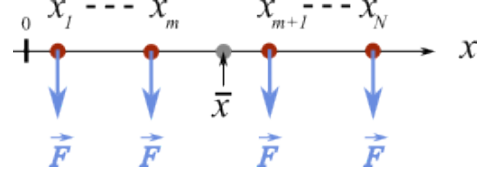


Figure 3: A set of N linearly vinculated point particles with the same mass m , represented respectively to the x -axis. The sought center of mass of this object is represented as \bar{x} . A total of m particles are at the left-hand side of \bar{x} , leaving $N - m$ particles at the right-hand side. Gravitation g is assumed to take place downwards.

Given that gravitation g is assumed to take place downwards, the same force $\vec{F} = m g$ will be induced at each of the identical point particles. It can be shown that no angular acceleration will be implied when the object is pivoted at position \bar{x} provided the following requirement is satisfied:

$$\tau_{clockwise} = \tau_{counterclock} \quad (3)$$

This condition leads to the following expression in the case of the situation depicted in Figure 3.

$$\begin{aligned} & -F(x_1 - \bar{x}) - \dots - F(x_m - \bar{x}) + \\ & + F(x_{m+1} - \bar{x}) + \dots + F(x_N - \bar{x}) = 0 \implies \\ \implies & \sum_{i=1}^m (x_i - \bar{x}) + \sum_{i=m+1}^N (x_i - \bar{x}) = 0 \implies \\ \implies & \sum_{i=1}^N (x_i - \bar{x}) = 0 \implies \boxed{\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i} \quad (4) \end{aligned}$$

As could be expected, the imposed torque symmetry led to the identification of the equilibrium position \bar{x} as corresponding to the *arithmetic mean* of the position of the point particles, therefore confirming the direct relationship between this type of mean and the physical concept of center of mass.

The extension of the above formulation to an object formed by a set of vinculated point particles with possibly distinct masses m_i , $i = 1, 2, \dots, N$, is relatively immediate by considering that the force \vec{F}_i induced at each particle i can be readily expressed as:

$$\vec{F}_i = m_i \vec{g} \quad (5)$$

Thus, it follows that:

$$\begin{aligned}
& -m_1 g(x_1 - \bar{x}) - \dots - m_m g(x_m - \bar{x}) + \\
& + m_{m+1} g(x_{m+1} - \bar{x}) + \dots + m_N g(x_N - \bar{x}) = 0 \implies \\
& \implies g \left(\sum_{i=1}^m m_i (x_i - \bar{x}) + \sum_{i=m+1}^N m_i (x_i - \bar{x}) \right) = 0 \implies \\
& \implies \sum_{i=1}^N m_i (x_i - \bar{x}) = 0 \implies \sum_{i=1}^N m_i x_i - \sum_{i=1}^N m_i \bar{x} = 0 \implies \\
& \implies \boxed{\bar{x} = \frac{1}{\sum_{i=1}^N m_i} \sum_{i=1}^N m_i x_i} \quad (6)
\end{aligned}$$

which corresponds to the weighted arithmetic mean of the values corresponding to the point mass positions x_i , weighted by the respective mass m_i .

The role of the center of mass in classical (Newtonian) mechanics is so special that, in addition to its direct relationship to the concepts of arithmetic and weighted arithmetic means, it also allows the summarization of some of the dynamic properties of a non-infinitesimal object while concentrating all its mass concentrated at the respective centroid (e.g. [10, 11, 12]).

5 Functionals, Statistics and Probability

Given some numeric data (discrete or continuous), such as a set S of numbers, it is possible to assign a respective scalar value to that set, which will henceforth be considered to be real. The mapping that implements this association is often called a *functional*.

A generic functional F can be mathematically expressed as:

$$F : f(x) \longrightarrow F(f(x)) \in \mathbb{R} \quad (7)$$

As already observed, the means of a set S of values are functionals on S . However, not every functional is a mean, including the mode, median, extremes and the range of S , to name but a few cases among infinite possibilities.

When considered from the perspective of statistics and probability (e.g. [2, 13, 3, 4, 5, 6, 7, 8]), the concept of mean relates directly to the *expectance* of a random variable X , which constitutes a functional that can be expressed in discrete form as:

$$E[X] = \sum_{i=1}^N x_i p_i(x) \quad (8)$$

where x_i are N the possible values of the discrete random variable X and p_i corresponds to the respective relative frequencies.

In case the discrete random variable is represented in terms of a set of N samples $S = \{s_1, s_2, \dots, s_N\}$, with $s_i \in \{x_1, x_2, \dots, x_N\}$, the respective arithmetic average can be alternatively *estimated* in terms of the respective average:

$$E[X] = \frac{1}{N} \sum_{i=1}^N s_i \quad (9)$$

A discussion of the relationship between Equations 8 and 9 can be found in [2].

In the case of a continuous random variable X , the expectance becomes:

$$E[X] = \int_{-\infty}^{\infty} x p(x) dx \quad (10)$$

where $p(x)$ is the probability density function of X .

It is also possible to define the *average* of a generic function $f(x)$, or *signal* (e.g. [14, 15, 6, 7, 8, 16]), within an interval $[a, b]$ as follows:

$$\bar{f}(x) = \frac{1}{b-a} \int_a^b f(x) dx \quad (11)$$

The concept of expectance can be readily generalized to the quantities known as *moments*, which are functionals defined as:

$$M_k[X] = E[X^k] = \int_{-\infty}^{\infty} x^k p(x) dx \quad (12)$$

where k is often taken as corresponding to a positive integer value.

Of particular interest are the *central moments*, which take into account a translation by the respective expected value $E[X]$, being defined as:

$$\begin{aligned}
\mathcal{M}_k[X] &= E[(X - E[X])^k] = \\
&= \int_{-\infty}^{\infty} (x - \mu_x)^k p(x) dx \quad (13)
\end{aligned}$$

The important quantity know as *variance* of a random variable X is a particular case of a central moment, obtained for $k = 2$:

$$\sigma_X^2 = \mathcal{M}_2[X] = \int_{-\infty}^{\infty} (x - \mu_x)^2 p(x) dx \quad (14)$$

The *standard deviation* of X can then be defined as:

$$\sigma_X = +\sqrt{\sigma_X^2} \quad (15)$$

6 Arithmetic Mean

The arithmetic mean of a set of sampled values s_i , $i = 1, 2, \dots, N$, can be defined as follows:

$$\mu = \frac{1}{N} \sum_{i=1}^N s_i \quad (16)$$

which, as seen in Section 5, corresponds to the *expectance* of the random variable as sampled by the discrete set of values x_i . In addition, as discussed in Section 4, the arithmetic mean of the coordinates of a set of points with the same mass yields the respective center of mass. As a consequence, the relative position of the arithmetic mean among a set of points becomes invariant to respective translations.

This quantity is also often referred to as the *average* of the values. Observe that while the average is a particular type of mean, there are several other means that are not averages.

As a simple example, the arithmetic mean of the set of values $S = \{1, 2, 3, 4, 5\}$ is:

$$\mu = \frac{1 + 2 + 3 + 4 + 5 + 6 + 7}{5} = 4$$

Figure 4 illustrates the arithmetic mean, as well as its comparison with other means considered in the present work, regarding the above data set (a) as well as its translated version (b), inclusion of an outlier point (c), as well as a logarithmically skewed version (d). Interestingly, the arithmetic mean resulted always in the middle of the other means in all these considered cases.

The arithmetic mean is always the largest among the three Pythagorean means. In addition, it is translational invariant to the values of the random variable and particularly sensitive to the presence of outliers.

7 Weighted Arithmetic Mean

The weighed arithmetic mean can be understood as an immediate generalization of the arithmetic mean taking place when *weights* are assigned to the respective random variables, as we have already discussed in Section 4 from the Newtonian mechanics point of view.

In the case of a set of observed values $x_i, i = 1, 2, \dots, N$ of a discrete random variable X , each with associated weight w_i , the *weighted arithmetic mean* can be expressed as:

$$\omega = \frac{1}{\sum_{i=1}^N w_i} \sum_{i=1}^N x_i w_i \quad (17)$$

In the case of a continuous random variable X in an interval $[a, b]$ with weights $w(x)$, we have:

$$\omega = \frac{1}{\int_a^b w_i dx} \int_a^b x w(x) dx \quad (18)$$

As a simple example of the calculation of the weighted arithmetic mean, we shall consider the following set describing the observed values x_i of a discrete random vari-

able X and their respective weights w_i in terms of respective tuples $[x_i, w_i]$:

$$S = \{[x_1 = 1, w_1 = 4]; [x_2 = 3, w_2 = 2]; [x_3 = 5, w_3 = 4]\}$$

The respective weighted arithmetic mean can be obtained from Equation 17 as:

$$\omega = \frac{1}{10} [(1)(4) + (3)(2) + (5)(4)] = 3$$

The weighted arithmetic mean shares several of its properties with the arithmetic mean from which it derives, including translational invariance and particular susceptibility to outliers.

8 Outliers and the Truncated Mean

Real-world data often incorporate *outliers* (e.g. [17, 18, 19, 20]), namely outcomes that are far too *far* and/or *unlikely* (i.e. have low probability), when compared with the other values in the same set of data.

Though there is not definitive manner of defining outliers, they can often be identified in terms of outcomes of a discrete random variable having probability smaller than a given threshold T . In the case of continuous variables, the probability can be estimated around a neighborhood of the respective values being considered as outcome.

The *truncated mean* typically corresponds to estimating any type of mean after a set of outliers has been discarded. The main motivation for this approach stems from the fact that it reduces the sensitivity of the means to the presence of outliers.

Figure 5 illustrates the sensitivity of the weighted arithmetic mean to outliers, and the effectivity of the respective truncated approach in reducing the susceptibility to outliers.

9 Geometric Mean

Given a set of *positive positive* values $x_i, i = 1, 2, \dots, N$, the respective *geometric mean* can be expressed as:

$$g = \sqrt[N]{\prod_{i=1}^N x_i} = \sqrt[N]{x_1 x_2 \dots x_N} \quad (19)$$

An interesting insight about particularly interesting applications of the geometrical mean can be obtained by taking the logarithm of both sides of the above expression, assuming $x_i > 0$, yielding:

$$\log g = \frac{1}{N} \sum_{i=1}^N \log x_i \quad (20)$$

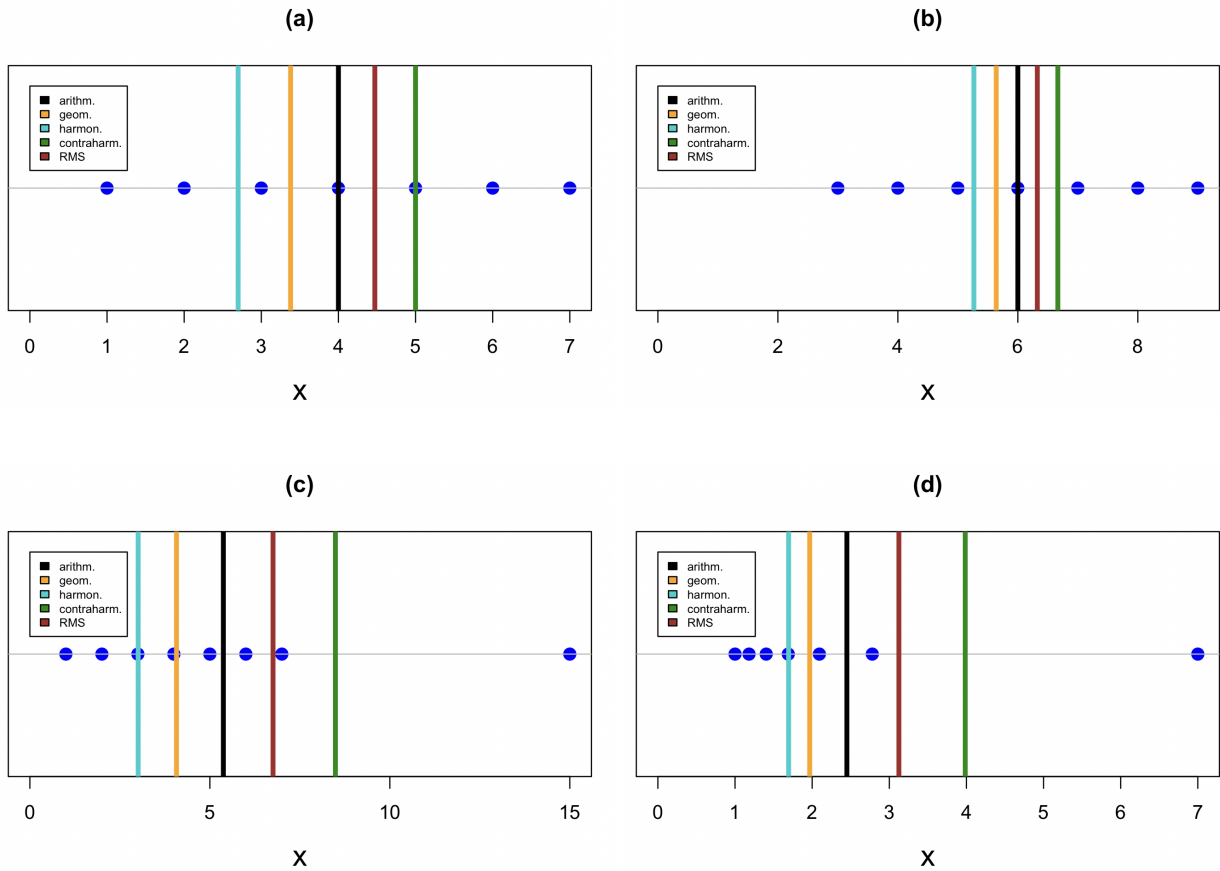


Figure 4: Illustration of the arithmetic, geometric, harmonic, and contra-harmonic means, as well as the root mean square, obtained for the set of values $S = \{1, 2, 3, 4, 5\}$ (a), as well as for its translation by 2 (b), incorporation of the outlier 15 (c), and after being logarithmically skewed (d). Observe the effect of the translation in approximating the means, as well as the relatively large influence on them by the other considered modifications.

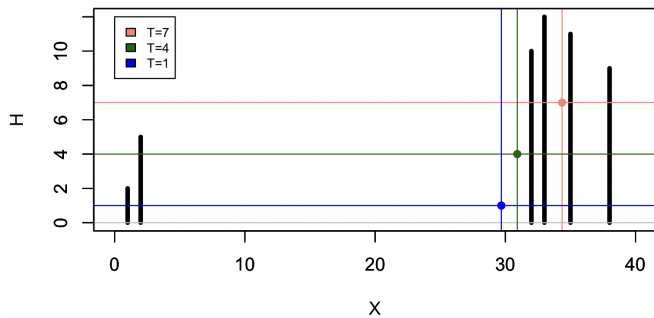


Figure 5: Illustration of the high sensitivity of the arithmetic mean respectively to outliers respectively to a discrete probability distribution of a discrete random variable X including two outlier values, namely 1 and 2. The number of observations of each value of X are shown as H . The average shown in blue is obtained in case all values are taken into account $T = 1$. By setting $T = 4$, the value $X = 1$ is discarded, resulting in a new respective truncated arithmetic mean show in green. The two outliers can be avoided by setting $T = 7$, which results in the truncated arithmetic mean shown in salmon. Observe that the three obtained means are greatly influenced by the presence of the outliers.

which corresponds precisely with the arithmetic mean of the set of values $\log x_i$. In other words, taking the geometric mean of the set x_i is the same as the logarithm of the arithmetic mean of the set $\log x_i$.

As a simple example of geometric mean calculation, consider the set of observations:

$$S = \{1, 2, 3, 4, 5, 6, 7\}$$

We then have that:

$$g = \sqrt[7]{7!} \approx 2.5598$$

A typical application of the geometric means is for summarizing the modification rate along M subsequent periods of time. For instance, let us say that the rate of increase of the height of a tree was observed to be as follows:

$$\begin{aligned} x_1 &= 0.05(5\%) && \text{along the first year;} \\ x_2 &= 0.01(1\%) && \text{along the second year;} \\ x_3 &= 0.07(7\%) && \text{along the third year.} \end{aligned}$$

A summarization of the overall growth rate may consider the respective geometric mean:

$$g = \sqrt[3]{(0.05)(0.01)(0.07)} \approx 0.032711$$

The summarization of the above growth rates in terms of the arithmetic mean would otherwise lead to:

$$\mu \approx 0.043333$$

The geometric mean, which is always the smallest among the three Pythagorean means, is often applied in economics.

10 Harmonic Mean

The *harmonic mean* of a set of N positive values x_i can be written as:

$$h = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}} = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_N}} \quad (21)$$

An interesting manner to better understand the harmonic mean is by rewriting the above expression as:

$$\frac{1}{h} = \frac{1}{N} \sum_{i=1}^N \frac{1}{x_i} \quad (22)$$

As a simple example of geometric mean calculation, consider the set of observations:

$$S = \{1, 2, 3, 4, 5, 6, 7\}$$

The respective harmonic mean can be obtained as:

$$h = \frac{7}{\frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{7}} \approx 2.69973$$

The harmonic mean, which is always the largest among the three Pythagorean means, is often applied in economics.

11 Lehmer and Contraharmonic Means

The *Lehmer mean* can be expressed as:

$$c = \frac{\frac{1}{N} \sum_{i=1}^N x_i^k}{\frac{1}{N} \sum_{i=1}^N x_i^{k-1}} \quad (23)$$

One of its particular cases (for $k = 2$) is the *contraharmonic mean*:

$$c = \frac{\frac{1}{N} \sum_{i=1}^N x_i^2}{\frac{1}{N} \sum_{i=1}^N x_i} = \frac{\frac{1}{N} (x_1^2 + x_2^2 + \dots + x_N^2)}{\frac{1}{N} (x_1 + x_2 + \dots + x_N)} \quad (24)$$

As a simple example of geometric mean calculation, consider the set of observations:

$$S = \{1, 2, 3, 4, 5, 6, 7\}$$

Leading to the following contraharmonic mean:

$$c = \frac{\frac{1}{7} (1 + 2 + \dots + 7)}{\frac{1}{7} (1^2 + 2^2 + \dots + 7^2)} = 5$$

The Lehmer and contraharmonic means are often considered in mathematics, especially in the areas of statistics and probability (e.g. [21, 22, 23]).

12 Root Mean Square – RMS

The *root mean square* of a set of N values x_i can be expressed as:

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} = \sqrt{\frac{1}{N} (x_1^2 + x_2^2 + \dots + x_N^2)} \quad (25)$$

In the case of a function $f(x)$ considered along an interval $[a, b]$, we have:

$$RMS = \sqrt{\frac{1}{b-a} \int_a^b f^2(x) dx} \quad (26)$$

One particularly interesting application of the RMS is for characterizing data sets and functions that have zero arithmetic average, of which the sinusoidal functions provide a prototypic example. By taking the observed values (scores) to the second power, the respective sum becomes necessarily positive, and therefore distinct from zero.

In the case of the sine and cosine, larger amplitudes will imply in larger RMS values. In particular, we have that the *RMS* along full periods of any of these two functions is equal to the amplitude multiplied by $\frac{1}{\sqrt{2}}$, as illustrated in Figure 6.

In electricity, the RMS value of an oscillating current or voltage wave corresponds to the value of a constant current or voltage that would dissipate the same thermal energy on a same electrical resistance R (e.g. [24, 25, 9, 26]).

As a simple example of geometric mean calculation, consider the set of observations:

$$S = \{1, 2, 3, 4, 5, 6, 7\}$$

The respective root mean square therefore is:

$$RMS = \sqrt{\frac{1}{7} (1^2 + 2^2 + 3^2 + \dots + 7^2)} \approx 4.47214$$

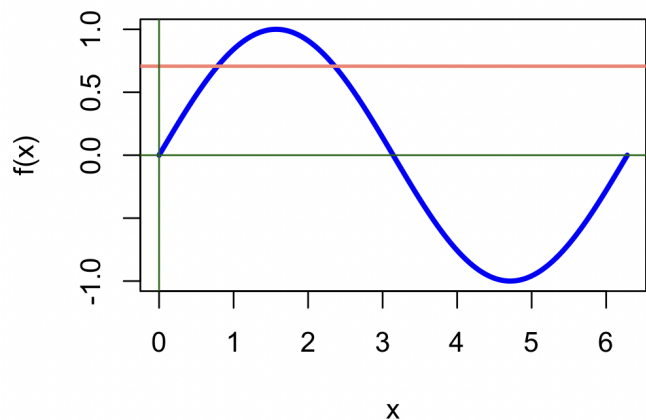


Figure 6: A full period of the sine function with amplitude 1 and its respective RMS value of $\frac{1}{\sqrt{2}}$ (in salmon).

13 Combining Means for Data Characterization

The fact that the types of means considered in this work are typically distinct one another suggests that they are providing mutually complementary information about the original data set, which can then be combined in order to obtain a set of *features* providing a more comprehensive description of the original data.

In this section we provide a brief illustration of this possibility, respective to three sets of 30 data elements sampled from the following probability density functions:

- (a) *uniform in the interval* [10, 20];
- (b) *normal with average 15 and variande 1;*
- (c) *normal with average 16 and variance 2.*

The above averages have been chosen large enough to avoid negative scores, which cannot be used in the geometric mean.

Figure 7(a) illustrates the respectively obtained scatterplot considering the respective arithmetic and geometric means of each of the 90 considered datasets. As could be expected, given the similar centers of mass and dispersions, the obtained clusters resulted particularly close one another.

By applying *principal component analysis* (PCA, e.g. [27, 28]) on the arithmetic, geometric, harmonic and contraharmonic means, as well as the root mean square characterization of the considered data elements, we obtain the projection shown in Figure 7(b), which presents a moderately wider separation between the groups.

The fact most of the variation of the original data have been accounted for by the two first principal components indicate that, at least for the specific set of data considered in this example, the respective means are highly

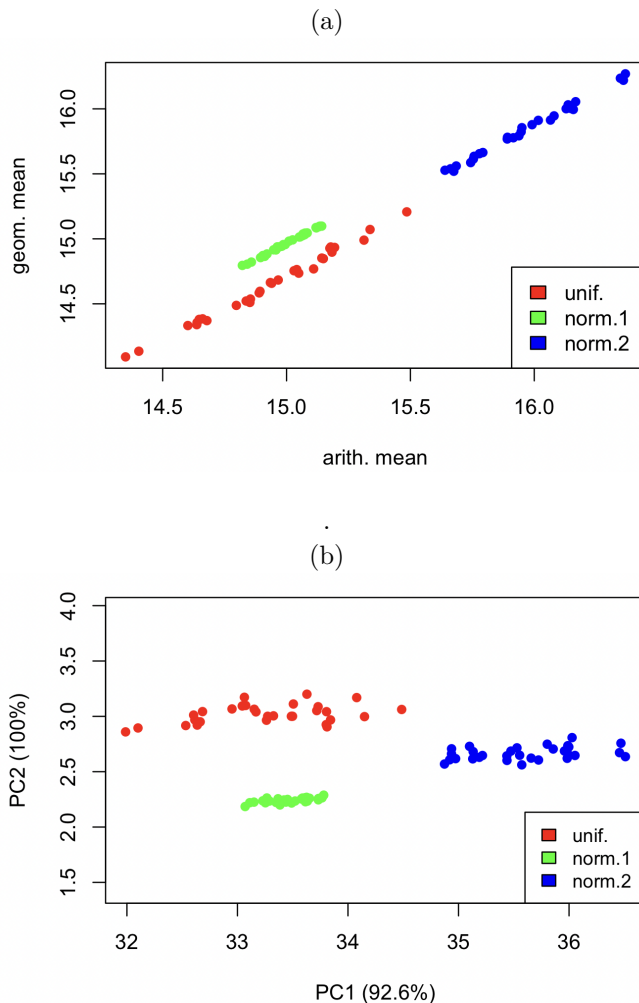


Figure 7: (a) Scatterplot between the arithmetic and geometric means obtained for the 3 considered datasets; and (b) two-dimensional projection, through principal component analysis (PCA), of the feature vectors incorporating the arithmetic, geometric, harmonic and contraharmonic means, as well as the root mean square obtained for the same datasets. The two axes correspond to the two principal components, whose percentage explanation of the variation of the original data is also respectively shown.

mutually correlated.

It should be observed that distinct tendencies can be possibly obtained for other types of data.

14 Concluding Remarks

Though not formally defined, the concept of mean plays a central role in virtually every area of science and technology. In addition to corresponding to a functional taking a set of values (or scores) into a respective scalar value, the several possible types of means also convey the idea of *centrality*, be it from the physical or statistical points of view, being also understood as implementing a *summa-*

rization of the values of the elements in the original data sets.

The present work aimed at presenting, illustrating and discussing several types of means in a hopefully accessible, though relatively abbreviated manner. The discussed types of means were: arithmetic, weighted, truncated, geometric, harmonic, Lehmer, contraharmonic and root mean square approaches.

After discussing possible interpretations of the meaning of the mean, as well as characterizing some of the main types of data typically available, the concept of mean was approached from the perspectives of Newtonian physics (center of mass) as well as functionals, probability and statistics (expectance).

In addition to providing respective presentation and illustration, the effects of data translation as well as the presence of outliers and skewness have also been briefly discussed subsequently. The possibility to combine several types of means while characterizing data has also been illustrated in terms of a simple specific example.

Acknowledgments

Luciano da F. Costa thanks CNPq (grant no. 307085/2018-0) and FAPESP (grant 15/22308-2).

Observations

As all other preprints by the author, the present work contains preliminary work subject to further revision and validation. Respective modification, commercial use, or distribution of any of its parts are not possible, as this work has author copyright. Many of the preprints by the author are also available in Hal and/or arXiv. This work can also be cited by using the DOI number or article identification link. Thanks for reading.

References

- [1] L. da F. Costa. Modeling: The human approach to science. Researchgate, 2019. https://www.researchgate.net/publication/333389500_Modeling_The_Human_Approach_to_Science_CDT-8. [Online; accessed 1-Oct-2020].
- [2] L. da F. Costa. Statistical modeling. https://www.researchgate.net/publication/334726352_Statistical_Modeling_CDT-13, 2019. [Online; accessed 10-Apr-2020].
- [3] N. Mukhopadhyay. *Probability and Statistical Inference*. CRC Press, New York, 2000.
- [4] M. H. DeGroot and M. J. Schervish. *Probability and Statistics*. Pearson, 2011.
- [5] E. Kreyszig. *Advanced Engineering Mathematics*. Wiley and Sons, 2015.
- [6] M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan. *Introduction to statistical signal processing with applications*. Prentice-Hall, Inc., 1995.
- [7] S. M. Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.
- [8] R. M. Gray and L. D. Davisson. *An introduction to statistical signal processing*. Cambridge University Press, 2004.
- [9] R. A. Serway and J. W. Jewett. *Physics for scientists and engineers*. Cengage learning, 2018.
- [10] D. Morin. *Introduction to classical mechanics: with problems and solutions*. Cambridge University Press, 2008.
- [11] S. T. Thornton and J. B. Marion. *Classical dynamics of particles and systems*. Cengage Learning, 2021.
- [12] J. V. José and E. Saletan. *Classical Dynamics: A Contemporary Approach*. Cambridge University Press, 1998.
- [13] L. da F. Costa. Randomness: A challenging central concept. https://www.researchgate.net/publication/371169292_Randomness_A_Challenging_Central_Concept, 2023.
- [14] E. O. Brigham. *Fast Fourier Transform and its Applications*. Pearson, 1988.
- [15] L. da F. Costa. Convolution! Researchgate, 2019. https://www.researchgate.net/publication/336601899_Convolution_CDT-14. [Online; accessed 09-March-2020].
- [16] M. Vetterli, J. Kovačević, and V. K. Goyal. *Foundations of signal processing*. Cambridge University Press, 2014.
- [17] L. da F. Costa. On counterexamples and outliers. https://www.researchgate.net/publication/348281283_On_Counterexamples_and_Outliers_CDT-51, 2021.
- [18] D. M. Hawkins. *Identification of outliers*, volume 11. Springer, 1980.

- [19] C. H. Comin, J. Tejada, M. P. Viana, A. C. Roque, and L. da F. Costa. Archetypes and outliers in the neuromorphological space. In *The Computing Dendrite*, pages 41–59. Springer, 2014.
- [20] D. Ghosh and A. Vogt. Outliers: An evaluation of methodologies. In *Joint statistical meetings*, volume 2012, 2012.
- [21] Y.-M. Chu and M.-K. Wang. Optimal lehmer mean bounds for the toader mean. *Results in Mathematics*, 61(3-4):223–229, 2012.
- [22] W. N. Anderson Jr, M. E. Mays, T. D. Morley, and G. E. Trapp. The contraharmonic mean of hsd matrices. *SIAM Journal on Algebraic Discrete Methods*, 8(4):674–682, 1987.
- [23] T.-H. Zhao, Y.-M. Chu, and B.-Y. Liu. Optimal bounds for neuman-sándor mean in terms of the convex combinations of harmonic, geometric, quadratic, and contraharmonic means. In *Abstract and Applied Analysis*, volume 2012. Hindawi, 2012.
- [24] P. E. Gray and C. L. Searle. *Electronic Principles: Physics, Models and Circuits*. John Wiley and Sons, 1969.
- [25] K. S. S. Kumar. *Electric Circuits and Networks*. Pearson Education India, 2009.
- [26] L. da F. Costa. Steady state analysis of AC circuits: An amalgamated approach. https://www.researchgate.net/publication/353739558_Steady_State_Analysis_of_AC_Circuits_An_Amalgamated_Approach_CDT-57, 2021. [Online; accessed 21-Aug-2021].
- [27] R. A. Johnson and D.W. Wichern. *Applied multivariate analysis*. Prentice Hall, 2002.
- [28] F. Gewers, G. R. Ferreira, H. F. Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. da F. Costa. Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys*, 54(4):1–34, 2021.