# MEAN$^{\text{ingful}}$ Functions

Jeffrey A. Barnett
(jbb@notatt.com)

## 1 Introduction

The term *mean* is certainly a popular descriptor that has been applied to a variety of real-valued functions that operate on real arguments. Various functions called means and related forms are extensively discussed in two classics of analysis, both named *Inequalities*: One by Hardy, Littlewood, and Polya [9] and the other by Beckenbach and Bellman [3]. Further, means are a recurring topic in probability and statistics texts, such as Feller [7], as well as the non-specialized mathematics literature, such as the *American Mathematics Monthly*, where dozens of related articles have appeared over the last few decades.

Since the term is in widespread use, any difficulty in finding a definition of means would be somewhat surprising. However, the Encyclopedic Dictionary of Mathematics prepared by the Mathematical Society of Japan [10] with 23 entries in its index does not offer one. The only definition that I could find was in Borwein and Borwein [4], where functions, $h\colon \Re^+ \times \Re^+ \to \Re^+$, are called means if $\min(x, y) \leq h(x, y) \leq \max(x, y)$ for all $x, y \in \Re^+$.

The intent of this article is to develop a reasonable definition of the term, mean, then discuss some general properties shared by those functions that satisfy this definition. My etymological approach has been to look for commonality in its various manifestations, particularly where a mean is used to estimate characteristic values of numerical measures attached to elements of a population. In fact, the definition offered herein might serve as a definition for a class of *estimator* functions except that that term has several connotations that are not intended.

The particular properties discussed below are conditions where evaluation of a mean can be decentralized (distributed) and a delineation of those means that can be represented as linear combinations of their arguments. It is also noted that the collection of means is not convex.

# 2 Definition

A mean is a real-valued function whose arguments are drawn from some connected subset of $\Re$, such as $\Re^+$. Thus, a mean is a function whose domain is $L^\star = L^1 \bigcup L^2 \cdots$, for some interval $L \subset \Re$, where $L^1 = L$ and $L^{n+1} = L \times L^n$ when $n \geq 1$. Thus, if $h$ is a mean, we are entitled to write $h\colon L^\star \to \Re$. Two common themes spotted in virtually all usage of the term are that means are symmetric and monotonic. A function, $h$, is *symmetric* in the sense used here, if

$$h(x_1, \ldots, x_n) \quad = \quad h(x_{i_1}, \ldots, x_{i_n})$$

for all permutations, $i_1, \ldots, i_n$, of the integers, $1, \ldots, n$, and it is *monotonic*, actually monotonic non-decreasing, if

$$h(x_1, \ldots, x_n) \quad \leq \quad h(y_1, \ldots, y_n),$$

whenever $x_i \leq y_i$ for $1 \leq i \leq n$. It also common that a mean have the *identity* property, that is to say $h(x, \ldots, x) = x$ for all $x \in L$.

So far, a mean could be any sequence, $h_1, h_2, \ldots$, of functions such that $h_n\colon L^n \to \Re$, for each $n \geq 1$, is symmetric, monotonic and has the identity property. As such, means lack a sense of coherence in their definition. The question is where and how to impose constraints *among* the $h_n$ that are consistent with the sort of examples we see when the rubric, mean, is used. The simplest such constraint is that it be *centralized*, i.e., that

$$h(x_1, \ldots, x_m) \;\leq\; h(x_1, \ldots, x_m, y_1, \ldots, y_n) \;\leq\; h(y_1, \ldots, y_n)$$

when $h(x_1, \ldots, x_m) \leq h(y_1, \ldots, y_n)$. Here, I am on terminological thin ice since I am unaware of centrality or any other name being previously dedicated to this property. Centrality has the virtue of applicability to many, if not most, functions that have been called means as well as being implicit in the sorts of usage that statisticians seem to have in mind when talking of means, medians, and the like; Statistical usage is to estimate various measures of populations and it would be decidedly strange if centralization were not required.

I am prepared to put a definition on the table to see if it agrees with your intuition as it does with mine.

**Definition 2.1** The function, $h\colon L^\star \to \Re$, where $L \subset \Re$ is a connected, non-degenerate interval and $L^\star = L^1 \bigcup L^2 \cdots$, is a *mean* if it has the identity property and is monotonic, symmetric, and centralized. Below, $h_n$, for $n \geq 1$ is used to denote $h$ restricted to $L^n$.

# 3 Examples and Otherwise

The Hölder means with $L = \Re^+$, are defined for each $c \in \Re \cup \{+\infty, -\infty\}$ as

$$H_c(a_1, \ldots, a_n) \quad = \quad \lim_{v \to c} \left[ \frac{1}{n} \sum_{i=1}^{n} a_i^v \right]^{\frac{1}{v}}$$

and are obvious examples of means. Thus, min and max are means as are the arithmetic, geometric, and harmonic means; the respective values of $c$ being $-\infty$, $+\infty$, 1, 0, and $-1$. The median, defined as the middle-magnitude element of an odd number of elements and the average of two middle-magnitude elements when there are an even number, is a mean. On the other hand, the mode is not because it isn't monotonic.

Another family of means, with $L = \Re$, is constructed, by choosing an arbitrary $r \in \Re$, as follows:

$$h_r(a_1, \ldots, a_n) \quad = \quad \begin{cases} \min(a_1, \ldots, a_n) & \text{if } \max(a_i) < r; \\ r & \text{if } \min(a_i) \leq r \leq \max(a_i); \\ \max(a_1, \ldots, a_n) & \text{if } r < \min(a_i); \end{cases}$$

This example shows that a mean need be neither continuous or homogeneous. The Gini means [8] are a family with two parameters, $u, v \in \Re$, defined as

$$G(a_1, \ldots, a_n; u, v) \quad = \quad \left( \sum_{i=1}^{n} a_i^{u+v} \Big/ \sum_{i=1}^{n} a_i^v \right)^{\frac{1}{u}} .$$

These functions are not monotonic for all values of $u$ and $v$ (Farnsworth and Orr [6]) so some are means, as defined herein, and some are not.

# 4 Basic Results

Several useful inferences about means are directly available from the above definition:

**Theorem 4.1** $\min(x_1, \ldots, x_n) \le h(x_1, \ldots, x_n) \le \max(x_1, \ldots, x_n)$.

**Theorem 4.2** Each $h_n$ is onto $L$, therefore, $h\colon L^\star \to L$.

**Theorem 4.3** $h(x_1, \ldots, x_m, y_1, \ldots, y_n) = h(x_1, \ldots, x_m)$ if $h(x_1, \ldots, x_m) = h(y_1, \ldots, y_n)$.

**Theorem 4.4** If $h$ is a continuous mean, then $h_n^{-1}(x) \subset L^n$ is a connected set for each $x \in L$ and each $n \ge 1$.

**Theorem 4.5** If $f\colon L \to L$ is a strictly monotonic continuous function and $h$ is a mean, then $h_f(x_1, \ldots, x_n) = f^{-1}(h(f(x_1), \ldots, f(x_n)))$ is a mean.

Theorem 4.1 follows from monotonicity and identity and Theorem 4.2 is a direct consequence. Theorem 4.3 follows from centrality and Theorem 4.4 follows from continuity, monotonicity, and the fact that the domain of each $h_n$ is convex. Theorem 4.5 is proved by showing that each defining property of a mean is preserved.

# 5 The Effects of Multiplicity

It is sometimes convenient to write the arguments to means as multisets. Thus, $h(A) = h(a_1, \ldots, a_n)$, where the $a_i$ are the $n = |A|$, not necessarily distinct, elements of the multiset $A$. In other words, the domain of $h$ is finite non-empty multisets formed from elements of $L$. This usage is sanctioned by the fact that $h$ is symmetric.

Next, the effect of multiplicity is examined. That venture is pursued by the following definition and theorem.

**Definition 5.1** If $h(A \bigcup \langle x \rangle) = h(A \bigcup \langle x, x \rangle)$ for all $A \in L^\star$ and $x \in L$, then the mean, $h$, is called an *extreme* mean. Angle brackets are used to emphasize that these are multisets.

Thus, an extreme mean is one that treats its argument as an ordinary set—one without multiplicity—since duplicates can be added/removed without changing the mean's value. The rational for calling these means extreme is

**Theorem 5.2** $h$ is extreme if and only if $h(A) = h(\min(A), \max(A))$.

"If" is straightforward as is "only if" when $|A| = 1$ or 2. So assume that $n > 2$ and choose an $A \in L^n$, where $A = \langle a_1 \leq \cdots \leq a_n \rangle$ and the $a_i$ are the $n = |A|$ elements of $A$ in a non-decreasing order. Then

$$\begin{aligned} h_{n-1}(a_1, a_3, \ldots, a_n) &= h_n(a_1, a_3, a_3, \ldots, a_n) \\ &\geq h_n(a_1, a_2, a_3, \ldots, a_n) \;=\; h(A) \\ &\geq h_n(a_1, a_1, a_3, \ldots, a_n) \\ &= h_{n-1}(a_1, a_3, \ldots, a_n) \end{aligned}$$

The first and last lines follow because duplicates can be added/removed without changing the value of $h$. The other lines follow because $h$ is monotonic. Therefore,

$$\begin{aligned} h(A \backslash a_2) &= h_{n-1}(a_1, a_3, \ldots, a_n) \\ &= h_n(a_1, a_2, a_3, \ldots, a_n) \\ &= h(A) \end{aligned}$$

where $A \backslash x$ is $A$ with $x$ removed. This procedure can be iterated $n - 2$ times to remove all but (one copy of) the minimum and maximum elements of $A$.

The next theorem provides another characterization of extreme means.

**Theorem 5.3** If $f \colon L^2 \to L$, is any monotonic non-decreasing function such that $f(x, x) = x$, then $h(A) = f(\min(A), \max(A))$ is an extreme mean.

Symmetry is immediate and monotonicity and identity are part of the definition, so it is only remains to show that $h$ is centralized, i.e., that $h(A) \leq h(A \bigcup B) \leq h(B)$ when $h(A) \leq h(B)$. Since $\min(A \bigcup B)$ is either $\min(A)$ or $\min(B)$ and $\max(A \bigcup B)$ is either $\max(A)$ or $\max(B)$, there are four cases to consider. Centrality follows in each case because $f$ is monotonic and Theorem 4.1 entails $\min(A) \leq \max(B)$.

# 6 Dominance

It is possible that one mean dominate another in the sense that $h(A) \leq g(A)$ for all $A \in L^\star$, e.g., the arithmetic mean is never less than the geometric mean. This suggests a related question: Can there be two means that always agree on the relative magnitudes assigned to multisets? In other words, are their distinct means, $h$ and $g$, such that $h(A) < h(B)$ if and only if $g(A) < g(B)$? The answer is no by an application of the calibration lemma:

**Lemma 6.1** Let $f_1$ and $f_2$ be surjections with common domain, $\mathcal{D}$, and range, $\mathcal{R} \subset \Re$, such that $f_1(x) < f_1(y)$ if and only if $f_2(x) < f_2(y)$. If for every $r \in \mathcal{R}$, there is a $c = c(r)$ such that $r = f_1(c) = f_2(c)$, then $f_1 \equiv f_2$.

The function $c(r) = \langle r \rangle$ plays the role of the calibration function for means; the identity property entails that $h(r) = g(r) = r$. Therefore, distinct means must disagree on the relative magnitudes of some multisets. Further, for every $m \geq 1$, there are multisets, $A, B \in L^m$, such that $g$ and $h$ disagree on the relative magnitudes of $A$ and $B$. Here, $c(r) = \langle m{\times}r \rangle$, where $\langle m{\times}r \rangle$ is the multiset that contains $m$ repetitions of the element, $r$, and nothing else.

# 7 Distributed Computation

In practice, means are used to calculate characteristic measures of populations. Sometimes those populations are quite large so gathering the sample at a central location can be difficult. Thus, we are motivated to seek methods to distribute the calculation rather than amass the data at a single site. Consider a simple example: Two agents collect the samples $A$ and $B$ in order to determine $\mathrm{avg}(A \bigcup B)$, where avg is the ordinary arithmetic mean. The first agent calculates $\mathrm{avg}(A)$ and the second calculates $\mathrm{avg}(B)$ then both pass their results to a central site where

$$\mathrm{avg}(A \bigcup B) \quad = \quad \frac{|A| \cdot \mathrm{avg}(A) + |B| \cdot \mathrm{avg}(B)}{|A| + |B|}$$

is calculated. This example motivates the following definition:

**Definition 7.1** The mean $h$ is *1-distributable* if there exist continuous functions, $d_n\colon L^n \to \Re$ and $d_{mn}\colon \Re \times \Re \to L$ for all $m, n \geq 1$, such that $h(A \bigcup B) = d_{mn}(d_m(A), d_n(B))$ for all $A, B \in L^\star$, where $m = |A|$ and $n = |B|$.

As a consequence of its definition, a 1-distributable mean must be continuous. Continuity avoids trivialities such as $d_m$ and $d_n$ each interleaving the digits of the decimal representations of their arguments and $d_{mn}$ reconstructing them.

If for a fixed $k$ and all $m, n > 1$ there are continuous $d_n\colon L^n \to \Re^k$ and $d_{mn}\colon \Re^k \times \Re^k \to L$, then $h$ is called *k-distributable*. Thus, all continuous extreme means are 2-distributable. Figure 1 shows the general layout of the distributed computation of a mean.

Another definition is necessary for the characterization of 1-distributable means:

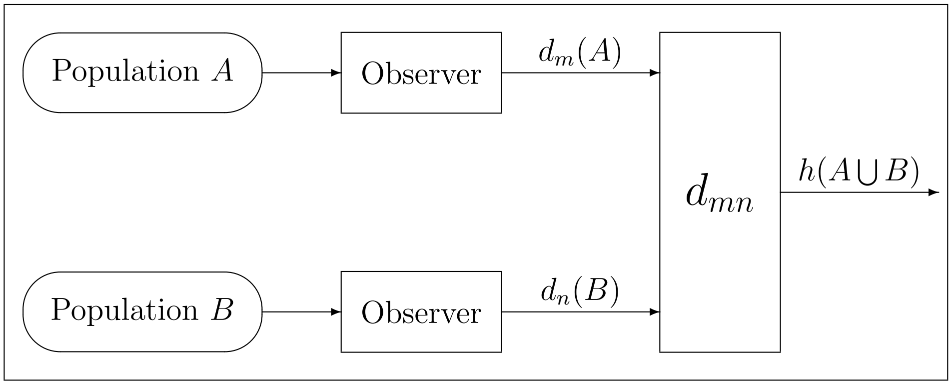


Figure 1: Distribute computation of *mean*.

**Definition 7.2** $A, B \in L^\star$ are *h-equivalent*, written $A \not\equiv B$, if $|A| = |B|$ and $h(A \bigcup C) = h(B \bigcup C)$ for all $C \in L^\star$.

Clearly $\not\equiv$ partitions $L^n$, for each $n \geq 1$, into equivalence classes and a necessary, but not sufficient, condition that $A$ and $B$ be in the same class is that $h(A) = h(B)$. The next theorem shows that the condition is sufficient as well when $h$ is 1-distributable.

**Theorem 7.3** The mean $h$ is 1-distributable if and only if (1) $h$ is continuous and (2) $A \not\equiv B$ whenever $|A| = |B|$ and $h(A) = h(B)$.

If $A \not\equiv B$ when $h(A) = h(B)$ and $|A| = |B|$, it is straightforward to select a unique element in each equivalence class induced by $\not\equiv$. A natural choice is $\langle m \times q \rangle$, where $m = |A|$, $q = h(A)$, and $\langle m \times q \rangle$ is the multiset that consists of exactly $m$ repetitions of $q$ and nothing else. Now define $d_m(A) = h(A)$, where $m = |A|$, and define $d_{mn}(x, y) = h(\langle m \times x \rangle \bigcup \langle n \times y \rangle)$ for each $m, n \geq 1$. The $d_m$ and $d_{mn}$ are continuous because $h$ is, so the "if" part of the theorem follows.

The "only if" part of the theorem will be proved if it can be shown that $h$ being 1-distributable implies $d_m(A) = d_m(B)$ when $h(A) = h(B)$ and $m = |A| = |B|$, because, for arbitrary $C \in L^\star$,

$$\begin{aligned} h(A \bigcup C) &= d_{mn}(d_m(A), d_n(C)) \\ &= d_{mn}(d_m(B), d_n(C)) \\ &= h(B \bigcup C), \end{aligned}$$

where $n = |C|$, and hence, $A \not\equiv B$. From these remarks, it is clear that the demonstration can be completed by showing, for all $m \geq 1$ and $A \in L^m$, that $d_m(A) = d_m(\langle m \times q \rangle)$, where $q = h(A)$, since the mean of every member of the equivalence class containing $\langle m \times q \rangle$ will have the same $d_m$ value.

Let $Z = \{\langle m \times x \rangle | x \in L\}$ and note that $d_m$ restricted to $Z$ must be a 1-to-1 strictly monotonic function since it is continuous. Assume that $d_m$ is increasing on $Z$ (otherwise, negate its values). To show that $d_m(A) = d_m(\langle m \times q \rangle)$ for any $A \in L^m$, where $q = h(A)$, let $r = \min(A)$ and $R = \max(A)$. If $r = R$ or $d_m(\langle m \times r \rangle) \leq d_m(A) \leq d_m(\langle m \times R \rangle)$, we are done. So assume that $d_m(\langle m \times R \rangle) < d_m(A)$. (The proof where $d_m(A) < d_m(\langle m \times r \rangle)$ is virtually identical.)

For each $1 \leq i \leq m$ and $0 \leq s \leq 1$, let $p_i(s) = (1-s)r + sa_i$ and define $P(s) = \langle p_1(s), \ldots, p_m(s) \rangle$. Hence, $P(0) = \langle m \times r \rangle$, $P(1) = A$, and $d_m(P(s))$ is continuous. By the intermediate value theorem, there must be a $0 < z < 1$ such that $d_m(P(z)) = d_m(\langle m \times R \rangle)$. Thus, $h(P(z)) = R$ but $p_i(z) < R$ for all $1 \leq i \leq m$, a contradiction (Theorem 4.1) since $h(P(z)) \leq \max(p_i(z))$. This completes the proof of Theorem 7.3.

The distribution theorem makes it immediate that the Hölder means, defined on $\Re^+$, are 1-distributable while the ordinary median is not. In fact, the ordinary median is not $k$-distributable for any fixed $k$ since its "storage complexity" is of the order $n/2$.

From theorems 7.3 and 4.4 we know that the equivalence classes induced by $\not\equiv$, where $h$ is 1-distributable, on $L^n$, for each $n \geq 1$, are connected sets. Further, there is a simple curve, namely $f_n(q) = \langle n \times q \rangle$, where $q \in L$, that intersects each equivalence class in exactly one point and all elements of an equivalence class can be continuously mapped onto that intersection. These observations motivate two conjectures about $k$-distributable means:

**Conjecture 7.4** If $h$ is a $k$-distributable mean, then for each $n \geq 1$, there is a continuous $f_n: L^k \to L^n$ such that $f_n(L^k)$ intersects each equivalence class induced by $\not\equiv$ on $L^n$ in exactly one point.

**Conjecture 7.5** If $h$ is a $k$-distributable mean, then there are continuous functions, $d_m: L^m \to \Re^k$ and $d_{mn}: \Re^k \times \Re^k \to L$ for all $m, n \geq 1$, where $d_{mn}(d_m(A), d_n(B)) = h(A \cup B)$, such that $d_m(C) = d_m(D)$ if and only if $C \not\equiv D$.

# 8 Convexity

It is straightforward to check that $\mathcal{E}_L$, the class of extreme means defined on $L^\star$, is convex because $h \in \mathcal{E}_L$ if and only if there exists a monotonic function, $f$, with the identity property such that $h(A) = f(\min(A), \max(A))$. Convexity follows since $f_z(x, y) = zf_1(x, y)+(1-z)f_2(x, y)$, where $0 \leq z \leq 1$, is monotonic and has the identity property whenever $f_1$ and $f_2$ do.

However, the class of all means, $\mathcal{H}_L$, with domain $L^\star$ is not convex as shown by an example: Since min and avg, the arithmetic average, are means $h(A) = .5 \cdot \min(A) + .5 \cdot \text{avg}(A)$ would necessarily be a mean if $\mathcal{H}_L$ were convex. But $h(3) = 3$, $h(1, 9) = 3$, and $h(3, 1, 9) = 8/3 \neq 3$ is a contradiction, Theorem 4.3, so $\mathcal{H}_L$ is not convex.

Another possibility is to liberalize the definition of convexity.

**Definition 8.1** The class of functions, $\mathcal{F}$, is $\mathcal{Q}$-*convex* if the function $f_q \in \mathcal{F}$ for all $q \in \mathcal{Q}$ and $f_1, f_2 \in \mathcal{F}$, where $f_q = q(f_1, f_2)$.

Ordinary convex is the case where $\mathcal{Q}$ is the collection of $q_z$, where $0 \leq z \leq 1$ and $q_z(f_1, f_2)(A) = zf_1(A)+(1-z)f_2(A)$. Of course every set of functions is $\mathcal{Q}$-convex when $\mathcal{Q} = \{I_1, I_2\}$, where $I_1(x, y) = x$ and $I_2(x, y) = y$ are identity functions. Barnett [2] shows that

**Theorem 8.2** If $\mathcal{H}_L$ is $\mathcal{Q}$-convex, then $\mathcal{Q} \subset \{I_1, I_2\}$.

Thus, it is not possible to define means as non-trivial combinations of a set of "boundary" functions. On the other hand, a simple characterization of all means formed from weighted sums of their arguments is available. These functions are the subject of Sections 9 and 10.

# 9 Generalized Medians

The ordinary median estimates the break-even point for a fair even-money over/under game: Player $P_U$ bets one unit that a single observation of a random variable will be less than the median while player $P_O$ bets one unit that the observation will be greater; the winner takes both bets. Wagers are returned when the observation is equal to the median.

Elements of a class of generalized medians estimate the break even points for fair over/under games for other betting odds, i.e., games where the two players bet different amounts. These medians are introduced below and shown to be means. The next section introduces a class of weighted-sum means and relates them to the class of generalized medians.

The generalized medians are defined in terms of the almost linear functions that approximate straight lines, through $(1,1)$ with slopes in $[0,1]$, as they might be rendered on a raster device such as a computer terminal.

**Definition 9.1** The function $\xi\colon I^+ \to I^+$ is *almost linear* if $\xi(1) = 1$ and $\xi(m)+\xi(n)-1 \leq \xi(m+n) \leq \xi(m)+\xi(n)$ for all $m, n \geq 1$. $\mathcal{J}$ is the collection of almost linear functions.

It is straightforward to show that $0 \leq \xi(n+1) - \xi(n) \leq 1$, hence, that $1 \leq \xi(n) \leq n$ for all $\xi \in \mathcal{J}$ and $n \geq 1$. Figure 2 shows the initial sequences

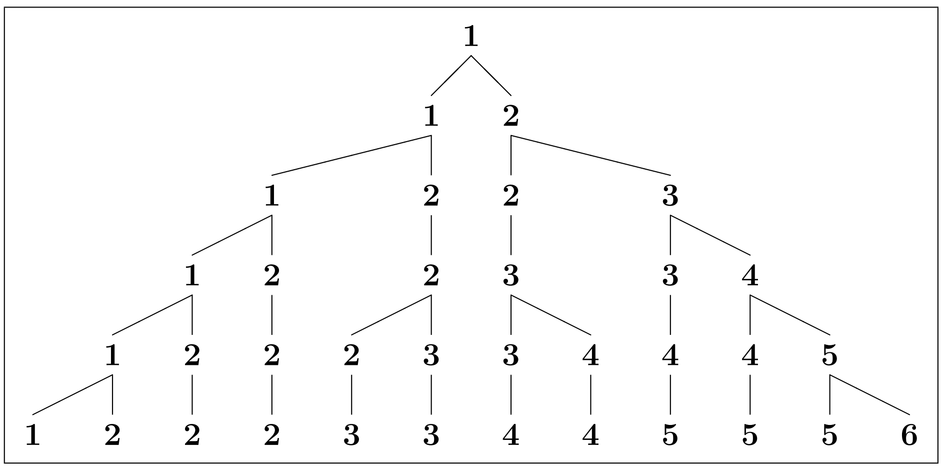


Figure 2: Initial sequences of $\xi(n)$, where $\xi \in \mathcal{J}$ and $n \leq 6$.

of the $\xi \in \mathcal{J}$ up to length six. It is interesting to note that the number of different initial sequences of length $n$ is $\sum_{i=1}^{n} \phi(i)$, where $\phi$ is the Euler totient function [1].

The classification of the $\xi \in \mathcal{J}$ provided by the next theorem simplifies the following discussion about generalized medians and weighted-sum means. Methods similar to those developed by Eisele and Hadeler [5] can be used to prove it.

**Theorem 9.2** $\mathcal{J}$ is partitioned into four classes:

1. If $\xi(n) = 1$ for all $n \geq 1$ or $\xi(n) = n$ for all $n \geq 1$, then $\xi$ is called *extreme*.

2. If $\xi(n) = \lfloor nr + 1 \rfloor = \lceil nr \rceil$ for all $n \geq 1$, where $0 < r < 1$ is irrational, then $\xi$ is called *irrational*.

3. If $\xi(n) = \lceil n\kappa/\pi \rceil = \lfloor n\kappa/\pi - 1/\pi \rfloor$ for all $n \geq 1$, where $0 < \kappa < \pi$ are integers and $(\kappa, \pi) = 1$, then $\xi$ is called *light*.

4. If $\xi(n) = \lfloor n\kappa/\pi + 1 \rfloor = \lceil n\kappa/\pi + 1/\pi \rceil$ for all $n \geq 1$, where $0 < \kappa < \pi$ are integers and $(\kappa, \pi) = 1$, then $\xi$ is called *heavy*.

If $\xi^-(n) = \lceil n\kappa/\pi \rceil$ and $\xi^+(n) = \lfloor n\kappa/\pi + 1 \rfloor$, then $\xi^-$ and $\xi^+$ are called a *rational pair*. Since $\kappa$ and $\pi$ are relatively prime, $\xi^+(n) = \xi^-(n)$ when $n \neq 0 \bmod \pi$ and $\xi^+(n) = \xi^-(n) + 1$ otherwise.

Next, median-like functions are defined using the $\xi \in \mathcal{J}$. That discussion is simplified by the following notation convention: Let $a_1 \leq \ldots \leq a_n$ be the $n = |A|$ elements of the multiset $A$ in non-decreasing order, then $A_i = a_i$.

**Definition 9.3** The following functions, $h\colon \Re^* \to \Re$, are *generalize medians*:

- If $0 \leq \lambda \leq 1$, then $h(A) = \lambda \min(A) + (1 - \lambda) \max(A)$ is an *extreme median*.

- If $\xi \in \mathcal{J}$ is irrational, then $h(A) = A_{\xi(|A|)}$ is an *irrational median*.

- If $\xi^+, \xi^- \in \mathcal{J}$ are a rational pair and $0 \leq \lambda \leq 1$, then $h(A) = \lambda A_{\xi^+(|A|)} + (1 - \lambda) A_{\xi^-(|A|)}$ is a *rational median*.

An extreme median is a linear combination of the multiset minimum and maximum and an irrational median simply selects one element of its multiset argument by its relative magnitude. A rational median also selects a single element when $|A| \neq 0 \bmod \pi$ or $\lambda = 0$ or $\lambda = 1$ but selects a linear combination of two adjacent elements when $|A| = 0 \bmod \pi$ and $0 < \lambda < 1$. The rational median with $\kappa/\pi = \lambda = 1/2$ is just the ordinary median.

**Theorem 9.4** The generalized medians are means.

Extreme medians are extreme means, Theorem 5.3, so they are clearly means. The proof that an irrational median is a mean is a simple counting argument. Let $h$ be an irrational median. It is easily seen to be symmetric and monotonic and have the identity property. To prove that it is centralized, we must show that $h(A) \leq h(A \bigcup B) \leq h(B)$ when $h(A) \leq h(B)$.

Let $m = |A|$, $n = |B|$, and $C = A \bigcup B$; Hence $h(A) = A_{\xi(m)}$, $h(B) = B_{\xi(n)}$, $|C| = m+n$, and $h(C) = C_{\xi(m+n)}$. Thus, $A_{\xi(m)} \leq B_{\xi(n)}$ by assumption. In $C$, there are at most $\xi(m) + \xi(n) - 2$ elements less than $h(A) = A_{\xi(m)}$: $A_1, \ldots, A_{\xi(m)-1}$ from $A$ and $B_1, \ldots, B_{\xi(n)-1}$ from $B$. Therefore, $C_i \geq h(A)$ for all $i \geq \xi(m)+\xi(n)-1$. Thus, $h(C) = C_{\xi(m+n)} \geq h(A)$ because $\xi(m+n) \geq \xi(m)+\xi(n)-1$ is a defining property of the almost linear functions. A similar argument establishes that $h(C) \leq h(B)$ and, hence, that $h$ is centralized.

The proof that a rational median, with $\lambda = 0$ or $\lambda = 1$, is identical to the above. More care with the counting argument is necessary when $0 < \lambda < 1$.

# 10 Weighted-Sum Means

Many recurring and important analytic forms are expressed as linear combinations of variables. If the form,

$$h(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i^n a_i,$$

where the $n$ in $w_i^n$ is a superscript, not an exponent, is a mean then

$$h(a_1, \ldots, a_n) = \left( \sum_{i=1}^{n} w_i^n a_i^z \right)^{\frac{1}{z}}$$

is also a mean (Theorem 4.5). However, considerations of symmetry and identity entail that $w_i^n = 1/n$ for all $1 \leq i \leq n$, so these forms are just Hölder means. A more inclusive set of forms is permitted by

**Definition 10.1** A *weighted-sum* mean, $h$, is a mean expressible as

$$h(A) = \sum_{i=1}^{n} w_i^n A_i, \quad \text{where } n = |A|.$$

The difference between this definition and the previous one, is that symmetry is exploited to order the elements of $A$ by magnitude.

**Theorem 10.2** The generalized medians are weighted-sum means.

Consider the three types of generalized medians in turn: When $h(A) = \lambda \min(A) + (1-\lambda) \max(A)$, then it is an extreme median. The corresponding $w_i^n$ are $w_1^1 = 1$ and $w_1^n = \lambda$ and $w_n^n = 1-\lambda$ for all $n > 1$; all other $w_i^n = 0$. The

$w_i^n$ for an irrational median are $w_{\xi(n)}^n = 1$ with all other $w_i^n = 0$. For the case of a rational median defined by the rational pair $\xi^-$ and $\xi^+$, with $(\kappa, \pi) = 1$ and $0 \leq \lambda \leq 1$, let $w_{\xi^-(n)}^n = \lambda$ and $w_{\xi^+(n)}^n = (1 - \lambda)$ when $n = 0 \bmod \pi$. When $n \neq 0 \bmod \pi$, let $\xi(n) = \xi^-(n) = \xi^+(n)$ and $w_{\xi(n)}^n = 1$. All other $w_i^n = 0$. This established the theorem.

A tedious argument about the possible locations of zeroes among the $w_i^n$ establishes the next theorem [2].

**Theorem 10.3** If $h$ is a weighted-sum mean, it is either the arithmetic average (where $w_i^n = 1/n$ for all $1 \leq i \leq n$) or a generalized median.

Therefore, the class of weighted-sum means, sans the arithmetic average, and the class of generalized medians are identical.

# 11 Discussion

The functions defined as means are meant to be used to estimate characteristic values of population measures from samples drawn from a parent population. That is the motivation for requiring symmetry, monotonicity, centrality, and the identity property. A less stringent requirement, that $h(A \bigcup B) = h(A)$ when $h(A) = h(B)$, is a plausible substitution for centrality. Both variations require sensible behavior when observations are combined.

Other requirements, e.g., homogeneity, differentiability, and continuity, can be levied for particular applications. For example, continuity was required in the discussion of distributed computations.

In the modern world, studies of which functional evaluations can be distributed and how are increasingly important as large computer ensembles become more prevalent. In that light, a characterization of $k$-distributable means is sought. Also a better characterization of the means might provide insight into the behavior of many venerable forms used in analysis.

# References

[1] J. A. Barnett, The number of almost-linear sequences, *American Mathematics Monthly*, Problems Section (Problem 1993, Solution 1996).

[2] J. A. Barnett, Primitive Estimation Functions, Northrop Research and Technology Center (1991). Postscript and TEX dvi files are available at `http://www.isi.edu/acmos/people/jeffb/jeffb.html`.

[3] E. F. Beckenbach and R. E. Bellman, *Inequalities*, Springer-Verlag (1965).

[4] J. M. Borwein and P. B. Borwein, *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*, John Wiley and Sons (1978).

[5] P. Eisele and K. P. Hadeler, Games of cards, dynamic systems, and a characterization of the floor and ceiling functions, *American Matematics Monthly* 97(6), pp. 466–477 (1990).

[6] D. Farnsworth and R. Orr, Gini Means, *American Mathematics Monthly* 93, (1986).

[7] W. Feller, *An Introduction to Probability Theorey and Its Applications*, John Wiley & Sons (1971).

[8] C. Gini, Di una formula comprensiva della medie, *Metron* 13 (1938).

[9] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge University Press (1959).

[10] Mathematical Society of Japan, *Encyclopedic Dictionary of Mathematics*, 3rd Edition, The MIT Press (1987).