



Received: 27 October 2015
Accepted: 26 January 2016
First Published: 10 February 2016

*Corresponding author:
Blane Hollingsworth, Scholar,
Mathematics Department, Indiana
University, Bloomington, IN, USA
E-mail: blane.hollingsworth@gmail.com.

Reviewing editor:
Nengxiang Ling, Hefei University of
Technology, China

Additional information is available at
the end of the article

STATISTICS | RESEARCH ARTICLE

The iteratile: A new measure of central tendency

Blane Hollingsworth^{1*}

Abstract: This article defines a new measure of central tendency (called the “iteratile”) by iterating a function which maps a triple of data into a triple of the median, mean, and trimean (sorted). An explicit formula is given with proof, along with brief discussion of potential applications.

Subjects: Applied Mathematics; Mathematics & Statistics; Science

Keywords: mean; median; outlier; central tendency

The ideas in DiMarco and Savitz (2013, 2012) take N values of data and study all their m -tiles, which separate the data in m sections, for $2 \leq m \leq N + 1$ (where the 2-tile is the median, the 4-tiles are quartiles, and the $(N + 1)$ -tile (the sample mean)). Observing that N values of data have N -many m -tiles, the author wondered if mapping all the data to their m -tiles and iterating would have a limit, and if so, what that limit would be.

In the case $N = 3$, we can get a formula. More precisely, given data (a, b, c) , for $a \leq b \leq c$, define $f^*(a, b, c) = (b, \frac{a+2b+c}{4}, \frac{a+b+c}{3})$, and then call $f(a, b, c)$ the sorted version of f^* , where the components are in increasing order. Denote f composed with itself n times by $f^{(n)}$.

THEOREM 1 $\lim_{n \rightarrow \infty} f^{(n)}(a, b, c)$ exists and is of the form (I, I, I) , where $I = \frac{3a+8b+3c}{14}$. (We call the limit the “iteratile” in what follows.)

Proof If $a = b = c$, we are done, so assume $a < b < c$.

Note first that the limit exists; f is a contraction mapping since each of b , $\frac{a+2b+c}{4}$, and $\frac{a+b+c}{3}$ must be greater than a and less than c .

We find the general formula with the aid of two lemmas. The first allows us to focus only on the behavior of the middle term, and the second shows by induction how the middle term exhibits geometric series behavior.

LEMMA 1 $f(a, b, c)$ takes either the form: $(b, \frac{a+2b+c}{4}, \frac{a+b+c}{3})$ or $(\frac{a+b+c}{3}, \frac{a+2b+c}{4}, b)$.

Proof By cases: Case 1: $b - a = c - b$.

The assumption is equivalent to $2b = a + c$, so $\frac{a+2b+c}{4} = \frac{4b}{4} = b$, and a similar argument yields $\frac{a+b+c}{3} = b$.

ABOUT THE AUTHOR

Blane Hollingsworth received his PhD from Auburn University in stochastic differential equations, and is currently a visiting scholar at Indiana University. Currently, he is studying alternatives to means and medians as measures of central tendency, especially for small sets of data with outliers/errors.

PUBLIC INTEREST STATEMENT

Often, small amounts of data and outliers make it difficult to do statistics. To measure the “typical” behavior, we usually use means or medians, but sometimes even those have problems. We offer an alternative way of measuring the center here.

Thus, $f(a, b, c) = (b, b, b)$, so $I = b$.

Case 2: $b - a > c - b$.

The assumption is equivalent to $2b > a + c$. Thus, $\frac{a+2b+c}{4} < \frac{2b+2b}{4} = b$ and $\frac{a+b+c}{3} < \frac{3b}{3} = b$. Further, $\frac{a+b+c}{3} - \frac{a+2b+c}{4} = \frac{a-2b+c}{12}$ which is negative by assumption. Thus,

$$b > \frac{a+2b+c}{4} > \frac{a+b+c}{3}.$$

Case 3: $b - a < c - b$. Similar arguments to Case 2 yield: □

$$b < \frac{a+2b+c}{4} < \frac{a+b+c}{3}.$$

Thus, it remains to study the difference from one iteration to the next in the middle term. Now, call $f^{(n)}(a, b, c) = (a_n, b_n, c_n)$ for convenience (so $a_0 = a$, etc.). We now give a formula for $b_{n+1} - b_n$ which will quickly complete the proof.

LEMMA 2 For any $n \in \mathbb{N}$, $b_{n+1} - b_n = \frac{a+c-2b}{4(-6)^n}$.

Proof By induction, we see that for $n = 0$:

$$b_1 - b_0 = \frac{a+2b+c}{4} - b = \frac{1}{4}(-a - c + 2b) = \frac{1}{4} \frac{a+c-2b}{(-6)^0}.$$

For $n = 1$, first observe by Lemma 1 that

$$b_2 = \frac{1}{4} \frac{5a+14b+5c}{6}, \tag{1}$$

since (without loss of generality) by Lemma 1,

$$b_2 = \frac{a_1+2b_1+c_1}{4} = \frac{1}{4} \left[\frac{a+b+c}{3} + \frac{a+2b-c}{4} + b \right] = \frac{1}{4} \frac{5a+14b+5c}{6}.$$

So,

$$b_2 - b_1 = \frac{1}{4} \frac{5a+14b+5c}{6} - \frac{a+2b+c}{4} = \frac{1}{4} \frac{a+c-2b}{(-6)^1}. \tag{2}$$

Now, suppose that $b_{n+1} - b_n = \frac{a+c-2b}{4(-6)^n}$ for an arbitrary $n \in \mathbb{N}$. We want to show that it holds for $n + 1$, so we study $b_{n+2} - b_{n+1}$.

An argument similar to (1) gives the relation:

$$b_{n+2} - b_{n+1} = \frac{1}{4} \frac{a_n+c_n-2b_n}{(-6)}. \tag{3}$$

So, Lemma 1 again gives

$$b_{n+2} - b_{n+1} = \frac{1}{4} [a_{n+1} - a_n + 2(b_{n+1} - b_n) + c_{n+1} - c_n], \tag{4}$$

and it seems we need to study $a_{n+1} - a_n$ and $c_{n+1} - c_n$ to continue.

Again, Lemma 1 tells us that $a_{n+1} - a_n$ will either be of the form $b_n - a_n$ or $\frac{a_n+b_n+c_n}{3} - a_n$, and similarly for $c_{n+1} - c_n$. Luckily, we need to only consider the sum of $a_{n+1} - a_n$ and $c_{n+1} - c_n$, so without loss of generality we know,

$$a_{n+1} - a_n + c_{n+1} - c_n = \frac{a_n + b_n + c_n}{3} + b_n - a_n - c_n = \frac{1}{6}[8b_n - 4a_n - 4c_n]. \quad (5)$$

Now we have what we need to prove the claim. By (4),

$$b_{n+2} - b_{n+1} = \frac{1}{4}[a_{n+1} - a_n + 2(b_{n+1} - b_n) + c_{n+1} - c_n],$$

and by (5) and induction hypothesis,

$$\begin{aligned} &= \frac{1}{4} \left[\frac{1}{6}(8b_n - 4a_n - 4c_n) + 2 \left(\frac{1}{4} \frac{a + c - 2b}{(-6)^n} \right) \right] \\ &= \left[\frac{1}{6}(2b_n - a_n - c_n) + 2 \left(\frac{1}{4} \frac{a + c - 2b}{(-6)^n} \right) \right]. \end{aligned}$$

Using (3), we get the equation

$$b_{n+2} - b_{n+1} = 4(b_{n+2} - b_{n+1}) + \frac{1}{2} \frac{a + c - 2b}{(-6)^n}.$$

We can write this as:

$$-3(b_{n+2} - b_{n+1}) = \frac{1}{2} \frac{a + c - 2b}{(-6)^n},$$

which reduces to

$$b_{n+2} - b_{n+1} = \frac{a + c - 2b}{(-6)^{n+1}},$$

and this proves the claim.

The upshot of Lemma 2 is that each subsequent iteration “steals $\frac{1}{6}$ from a (and from c) and gives it to b ,” which means we can set up the equation

$$\sum_{n=0}^{\infty} b_{n+1} - b_n = \sum_{n=0}^{\infty} \frac{a + c - 2b}{4(-6)^n}.$$

The left-hand side is a telescoping series, and we know that $\lim_{n \rightarrow \infty} b_n = I$. Also, the right-hand side is a simple geometric series, thus the above becomes

$$I - b = \frac{3}{14}(a + c - 2b),$$

or

$$I = \frac{1}{14}(3a + 8b + 3c),$$

as claimed. □

Can this be extended to higher n ? Unfortunately, there does not exist a universally agreed-upon notion of “percentile”; different definitions will yield different iteratiles. Thus, without specifying exactly which notion is desired, one cannot expect a general result. It seems that one gets a “symmetric convex combination” in any case.

How “good” or “useful” is the iteratile? One potential use is when there are errors in the data, so the data-set changes. What measure of central tendency minimizes the difference? Results along

these lines are discussed in DiMarco, Hollingsworth, and Savitz (2015) in the context of z-scores and normal distributions.

For example, if the data-set (0, 0, 24) were changed to (0, 3, 16), then the difference in the iteratiles would be zero. But of course, one is not allowed to see the data change before selecting the appropriate measure. How does one pick a central tendency measure a priori to minimize the difference in measure for an arbitrary data change? There is a general trend: when data in the middle is (expected to be) altered, use the mean, and when data at the extremes are altered, use the median. When there is an “inbetween” for what to use? At least, relative to the trimean, the iteratiles should be used if values toward the extreme are altered more, and the trimean when values toward the middle are altered more. Also, it seems that skewed alterations suggest use of iteratiles.

In general, it is quite crude to have to resort to the median, as so much information is unaccounted for. The iteratiles give a bit more of that information and reduce the effects of outliers. As one piece of data gets arbitrarily large, the non-median measures will go to infinity, but at least, with different speeds (mean: 1/3, trimean: 1/4, iteratiles: 3/14). As a result, perhaps using the iteratiles is a decent alternative to the median. But still, is there anything special about the iteratiles? One could, for example, merely make a measure of central tendency like $\frac{3a+10b+3c}{16}$, which would do even better ... but would it be optimal? Is there really something coming from the limit? These are complicated, long-standing, and potentially unanswerable questions, and even if the iteratiles itself turns out to be inferior, perhaps the idea of iteratiles may lead to other more useful alternatives and generalizations. At the very least, those who appreciate calculus may simply enjoy the result at a purely esthetic level.

Funding

The author received no direct funding for this research.

Author details

Blane Hollingsworth¹

E-mail: blane.hollingsworth@gmail.com

¹ Mathematics Department, Indiana University, Bloomington, IN, USA.

Citation information

Cite this article as: The iteratiles: A new measure of central tendency, Blane Hollingsworth, *Cogent Mathematics* (2016), 3: 1149135.

References

- DiMarco, D., & Savitz, R. (2012). The M-tile means, A new class of measures of central tendency: Theory and applications. *JIMS*, 12, 48–56.
- DiMarco, D., & Savitz, R. (2013). The M-tile deviation: A new class of measures of dispersion. *International Journal of Business Research*, 13, 117–124.
- DiMarco, D., Hollingsworth, B., & Savitz, R. (2015). On resistant versions of the standard score. *European Journal of Marketing*, 15, 7–16.



© 2016 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

