Efficient Weighted Average Calculation Algorithm

Matthew Tambiah

FAST Weighted Average Calculation Algorithm

Matthew A. Tambiah
Bachelor's in Electrical and Computer Engineering
Harvard University, Cambridge, United States of America
Founder and Chief Education Officer of Fast Education Group
https://orcid.org/0000-0001-6761-3723

Abstract

This paper presents an efficient algorithm (method) for calculating Weighted Averages called the "*Fast* Weighted Average" calculation algorithm. In general, the algorithm presented in textbooks for calculating a Weighted Average of $\boldsymbol{n}$ elements (each with a Values and a Weight) will require $\boldsymbol{n}$ multiplication operations, assuming no Weights and no Values are zero.

The Fast Weighted Average calculation algorithm can calculate the Weighted Average of $\boldsymbol{n}$ elements using **at most** $(\boldsymbol{n-1})$ multiplication operations when using Normalized Weights. Furthermore, in many examples the computational complexity of the multiplication operations of the *Fast* Weight Average algorithm are reduced compared to the standard Weighted Average calculation algorithm, as the multiplication operations of the Fast Weighted Average algorithm often have fewer digits, though this depends on the data set. The *Fast* Weighted Average algorithm is particularly useful when *people* perform and interpret the calculations and make decisions based on the results.

Review

The Weighted Average of $\boldsymbol{n}$ data points or Values of variable $\boldsymbol{x}$ with Values $\{x_1,\ x_2,\ \ldots\ x_i,\ \ldots\ x_n\}$, each with corresponding **non-negative** Normalized Weights $\{w_1,\ w_2,\ \ldots\ w_i,\ \ldots\ w_n\}$ is denoted by $\boldsymbol{\bar{x}}$ and calculated as:

$$\text{Weighted Average of } x\text{:} \qquad \bar{x} = \sum_{i=1}^{n} (w_i \cdot x_i)$$

Normalized Weights mean that:

$$\sum_{i=1}^{n} w_i = 1$$

If arbitrary weights are given $\{W_1,\ W_2,\ \ldots\ W_i,\ \ldots\ W_n\}$, which must be **non-negative,** they can be normalized as follows where $\boldsymbol{w_i}$ represents the Normalized Weight:

$$w_i = \frac{W_i}{\sum_{i=1}^{n} W_i}$$

I will assume none of the $\boldsymbol{w_i}$ values are 0. If any are zero, then they will ignored and excluded from the count of $\boldsymbol{n}$. So we will assume $\boldsymbol{n}$ positive Normalized weights $\{w_1, w_2, \ \ldots \ w_i, \ \ldots \ w_n\}$.

Standard implementation of the Weighted Average calculation with Normalized Weights of $\boldsymbol{n}$ elements will require $\boldsymbol{n}$ multiplication operations if each Weight is positive (not zero) and each value is non-zero. If any of the values are zero, then this reduces the number of multiplication operations required, but standard algorithms may require up to $\boldsymbol{n}$ multiplication operations.

In practice, any Values where either the value is zero or the weight is zero can be discarded (after Weights are Normalized, if needed) and the number of elements and multiplication operations reduced accordingly. This paper deals with situations where Weights are Normalized and each of the $\boldsymbol{n}$ elements in the Weighted Average calculation have positive (non zero, non-negative) Weights, and non-zero Values.

FAST Weighted Average Calculation Algorithm

The Fast Weighted Average calculation algorithm is implemented as follows. First, a value in the set of variables is selected as a Reference Value. We will use $\boldsymbol{x_1}$ as the Reference Value for the purposes of the formula. Any Value can be used as the Reference Value, and can simply be relabeled as $\boldsymbol{x_1}$ without loss of generality.

Define $\boldsymbol{\Delta x_i}$ as follows:

$$\Delta x_i = x_i - \boldsymbol{x_1} \qquad \text{(Difference between } x_i \text{ and the Reference Vale of } x_1\text{)}$$

This means:

$$x_i = \boldsymbol{x_1} + \Delta x_i \qquad [\text{Note that } \Delta x_1 = 0]$$

The *Fast* Weighted Average calculation algorithm to calculate the Weighted Average of $\boldsymbol{x}$, denoted by $\boldsymbol{\bar{x}}$, is expressed in the following formula:

$$\bar{x} = x_1 + \sum_{i=\mathbf{2}}^{n} (w_i \cdot \Delta x_i)$$

Note that the index of the sum $\boldsymbol{i}$ goes from **2** to $\boldsymbol{n}$, and so this requires $(\boldsymbol{n-1})$ multiplication operations. The sum would run from $\boldsymbol{i}$ going from **1** to $\boldsymbol{n}$, but the first term is zero since $(\Delta x_1 = 0)$, so the term when $(\boldsymbol{i=1})$ can be ignored.

I will call the terms $\boldsymbol{\Delta x_i}$ is the **Difference** (from a Reference Value) and each term $(\boldsymbol{w_i \cdot \Delta x_i})$ is a **Weighted Difference** term.

Note that the Difference terms are generally smaller than the original values if the values are close to each, so this method can be easier for people to execute with pen and paper or calculators and easier for people to estimate the Weighted Average given the values and Normalized Weights.

FAST Weighted Average Algorithm:

1) Discard any values which have zero Normalized weight

2) Select a Reference Value from the set of data points, which is referred to ax x_1

3) Calculate the Difference between every other value and the Reverence Value

4) Calculated the Weighted Difference of each term (aside from the Reference Value) by multiplying the Weight times the Difference from the Reference Value.

5) Sum the Weighted Differences

6) Add the Sum of the Weighted Differences to the Reference Value of x_1

Arbitrary Reference Value

The formula can be modified if different values of $\boldsymbol{x}$ are selected for the Reference Value or an Arbitrary value is selected as a Reference Value. Define the Reference Value as $\boldsymbol{A}$ ("**A**" is for **A**nchor), which can be **any value**.

Now define $\boldsymbol{\Delta x_i}$ as follows:

$$\Delta x_i = x_i - \boldsymbol{A}$$ (difference between x_i and the Reference Value or Anchor)

The Weighted Average of $\boldsymbol{x}$ can be calculated as:

$$\bar{x} = A + \sum_{i=1}^{n} (w_i \cdot \Delta x_i)$$

In general, this formula will require $\boldsymbol{n}$ multiplication operations. However, if the Reference Value $\boldsymbol{A}$ is selected to be $\boldsymbol{x_j}$, meaning $(\boldsymbol{A} = \boldsymbol{x_j})$, then $(\boldsymbol{\Delta x_j} = \boldsymbol{0})$ and the corresponding Weighted Difference term can excluded from the sum so only $(\boldsymbol{n} - \boldsymbol{1})$ multiplication operations are required.

That is, if $(j = 1)$ and $(A = x_1)$ the sum runs from **2** to $\boldsymbol{n}$ as given in the first instance:

$$\bar{x} = A + \sum_{i=\boldsymbol{2}}^{n} (w_i \cdot \Delta x_i)$$

If $(j = n)$ and $(A = x_n)$the sum runs from **1** to ($\boldsymbol{n}$ – **1**):

$$\bar{x} = A + \sum_{i=1}^{\boldsymbol{n-1}} (w_i \cdot \Delta x_i)$$

If $(1 < j < n)$ the sum excludes $\boldsymbol{j}$:

$$\bar{x} = A + \sum_{i=1}^{\boldsymbol{j-1}} (w_i \cdot \Delta x_i) + \sum_{i=\boldsymbol{j+1}}^{n} (w_i \cdot \Delta x_i)$$

In each of these cases, there are only $(\boldsymbol{n-1})$ multiplication operations required if $(\boldsymbol{A = x_j})$ for some $\boldsymbol{j}$ rather than $\boldsymbol{n}$ multiplication operations with the traditional Weighted Average formula.

In addition to reducing the number of multiplication operations, the Weighted Difference algorithm is generally easier for people to calculate as the Difference terms are smaller and often have fewer digits than the original Values when the Values are close together so the computational complexity of the multiplication operations is reduced because there are fewer digits to multiply. This applies in many practical circumstances but may not be applicable in certain circumstances.

The remainder of the document is designed for a wide audience — therefore, it may use slightly different terms for concepts than are used in academic mathematics papers. For example, "Normalized Weights" are sometimes called "Percentage Frequencies".

Expected Value

Also note that the Expected Value calculation is essentially the same calculation as the Weighted Average calculation where instead of using the frequency percentage, $\boldsymbol{p_i}$ is the Probability of the given event occurring, and $\boldsymbol{N_i}$ is the Value assigned to that event.

The Expected Value is essentially the Weighted Average outcome of the Value based on the probabilities. The Expected Value of $\boldsymbol{N}$ is denoted as $\boldsymbol{\overline{N}}$ and is calculated as:

$$\text{Expected Value of } N\text{:} \qquad \overline{N} = \sum_{i=1}^{n} (p_i \cdot N_i)$$

In general, calculating the Expected Value will require $\boldsymbol{n}$ multiplication operations for $\boldsymbol{n}$ values which have non-zero (and positive) probabilities and when none of the values are zero. The calculation can be simplified by selecting a Reference Value (or Anchor) equal to one of the values (N_j for some j) similar to the Weighted Average. Everything stated about Weighted Average calculations applies to the Expected Value. So an Expected Value of $\boldsymbol{n}$ outcomes can be calculated with at most $(\boldsymbol{n-1})$ multiplication operations if the probabilities and Values associated with each outcome are known.

Weighted Average (or Expected Value) of Two Values

Given two values $(N_1,\ N_2)$ we can simply the Weighted Average calculation. Note that the Weighted Average of two Values will always be between the two Values, which should provide some intuition regarding whether to add or subtract the Weighted Difference from the Reference Value. I also use the term Percentage Frequency for Normalized Weight and use p_i instead of w_i. Given two Values $(N_1,\ N_2)$ and corresponding Percentage Frequencies of $(p_1,\ p_2)$ (Normalized Weights):

$$\overline{N} = \sum_{i=1}^{2} (p_i \cdot N_i) = (p_1 \cdot N_1) + (p_2 \cdot N_2)$$

if either $\boldsymbol{p_1}$ or $\boldsymbol{p_2}$ is zero, then no multiplication is required as then either $\boldsymbol{p_1}$ or $\boldsymbol{p_2}$ is will be 1.

Without loss of generality, we can say $(N_2 \geq N_1)$ as we simply label the larger of two values $\boldsymbol{N_2}$. Define the variable $\boldsymbol{D}$ as Difference between $\boldsymbol{N_1}$ and $\boldsymbol{N_2}$:

$$D = N_2 - N_1 \qquad (D \text{ will be positive if } N_2 \geq N_1)$$

Note that $N_2 = N_1 + D$

Thus, by the definition, $\boldsymbol{D}$ will be positive. The Weighted Average of $(N_1,\ N_2)$, which is written as $(\overline{N})$ can be calculated as either:

$$\overline{N} = N_1 + (D \cdot p_2) \qquad \text{(Version 1)}$$

$$\overline{N} = N_1 - (D \cdot p_1) \qquad \text{(Version 2)}$$

As discussed earlier, these equations only require a **single** multiplication operation to execute. Do you understand why this works? What does your intuition tell you?

Average of Two Numbers

Let's consider the example where we have to calculate the Unweighted (or Direct) Average of two values/numbers $(N_1,\ N_2)$. This is also equivalent to the scenario where these values have equal Weight in a Weighted Average. In this section $(\overline{N})$ will refer to the Straight (Unweighted or Equally Weighted) Average.

The Traditional formula for the Average is:

$$\overline{N} = \frac{N_1 + N_2}{2}$$

If the two values are the same, then the Average is just the value which is repeated, and no further calculation is required: if $N_1 = N_2$ then $\overline{N} = N_1 = N_2$.

If the two values $(N_1,\ N_2)$ are different, let $\boldsymbol{N_2}$ be the larger of the two values. We can do this without loss of generality.

Define the Difference $\boldsymbol{D}$ as:

$$D = N_2 - N_1$$

Define $\boldsymbol{d}$ as:

$$d = \frac{D}{2}$$

Note that both $\boldsymbol{D}$ and $\boldsymbol{d}$ are positive based on $\boldsymbol{N_2}$ be the larger of the two values.The Average can then be calculated as:

$$\overline{N} = N_1 + d = N_1 + \frac{D}{2} \Longrightarrow N_1 = \overline{N} - d$$

$$\overline{N} = N_2 - d = N_2 - \frac{D}{2} \Longrightarrow N_2 = \overline{N} + d$$

The average of two numbers can also be thought of as the **Midpoint** of two numbers, and many Algebra classes teach a Midpoint formula equivalent to the average formula above. In this case, $\boldsymbol{M}$ is denoted as the Midpoint of two values $(N_1,\ N_2)$ and the "Traditional" Midpoint formula is:

$$M = \frac{N_1 + N_2}{2}$$

Therefore:

$$M = N_1 + d = N_1\ + \frac{D}{2} \Longrightarrow N_1 = M - d$$

$$M = N_2 - d = N_2\ - \frac{D}{2} \Longrightarrow N_2 = M + d$$

In this context, the Average or Midpoint is the point that is in the Middle of the two values. That is, the "Distance" or "Difference" from the Midpoint to each of the values N_1 and N_2 is **the same**. In this example, the "Difference" really means the Absolute Value of the Difference. The idea is that form the Midpoint, you must travel the same distance (in opposite directions) to reach N_1 and N_2.

To summarize:

$$M = \overline{N} = N_1 + d = N_1\ + \frac{D}{2}$$

$$M = \overline{N} = N_2 - d = N_2\ - \frac{D}{2}$$

Proof of *FAST* Weighted Average Algorithm

In general, a Weighted Average calculation of $\boldsymbol{n}$ values can be done with $(\boldsymbol{n-1})$ multiplications, given Normalized Weights.

As discussed, Weighted Average of N is denoted by $\overline{N}$ and calculated by:

$$\overline{N} = \sum_{i=1}^{n} (p_i \cdot N_i)$$

Where $\boldsymbol{p_i}$ is the Percentage Frequency or Normalized Weight for each value ($\boldsymbol{w_i}$ in previous sections). Now we will define each value relative to $\boldsymbol{N_1}$ as a **Reference Value**. Define $\boldsymbol{\Delta N_i}$ as:

$$\Delta N_i = N_i - \boldsymbol{N_1} \qquad \text{[We subtract } N_1 \text{ from each value]}$$

This means:

$$N_i = \boldsymbol{N_1} + \Delta N_i$$

Inserting this expressions for $\boldsymbol{N_i}$ into the Weighted Average Formula:

$$\overline{N} = \sum_{i=1}^{n} (p_i \cdot N_i) = \sum_{i=1}^{n} \left(p_i \cdot (N_1 + \Delta N_i)\ \right)$$

Multiplying through by $\boldsymbol{p_1}$:

$$\overline{N} = \sum_{i=1}^{n} \left(\ (p_i \cdot N_1) + (p_i \cdot \Delta N_i)\ \right)$$

Separating terms of the sum:

$$\overline{N} = \sum_{i=1}^{n} (p_i \cdot N_1) + \sum_{i=1}^{n} (p_i \cdot \Delta N_i)$$

$\boldsymbol{N_1}$ is a constant value which can be factored out:

$$\overline{N} = N_1 \cdot \sum_{i=1}^{n} (p_i) + \sum_{i=1}^{n} (p_i \cdot \Delta N_i)$$

We also have the constraint that:

$$\sum_{i=1}^{n} p_i = 1$$

as the Percentage Frequencies (Normalized Weights) must sum to 100%. Therefore, the last expression becomes:

$$\overline{N} = N_1 \cdot 1 + \sum_{i=1}^{n} (p_i \cdot \Delta N_i)$$

This simplifies to our ***Fast* Weighted Average Formula** for the Weighted Average Calculation:

$$\overline{N} = N_1 + \sum_{i=1}^{n} (p_i \cdot \Delta N_i)$$

Also note that in this context

$$\Delta N_1 = 0 \qquad \text{as} \qquad \Delta N_1 = N_1 - N_1 = 0$$

Therefore, the term for $(i = 1)$ will be zero, so the index can run from **2** to $\boldsymbol{n}$ which saves one multiplication operation.

$$\overline{N} = N_1 + \sum_{i=\mathbf{2}}^{n} (p_i \cdot \Delta N_i)$$

We can pick whichever value we want to serve as the Reference value by assigning it the label N_1. We therefore have freedom over which value we use as a Reference Value.

Proof of Arbitrary Reference Value Formula

We can also perform Weighted Average calculations with an arbitrary Reference Value. Let's define the reference value with the variable $\boldsymbol{A}$ for **Anchor**.

Now we define all our values relative to the Anchor $\boldsymbol{A}$:

$$N_i = A + \Delta N_i \qquad \text{where} \qquad \Delta N_i = N_i - A$$

We can go through the prior algebra where ether only difference is $\boldsymbol{N_1}$ becomes $\boldsymbol{A}$.

$$\overline{N} = \sum_{i=1}^{n} (p_i \cdot N_i) = \sum_{i=1}^{n} \big(p_i \cdot (A + \Delta N_i)\ \big)$$

Multiplying through by $\boldsymbol{p_1}$:

$$\overline{N} = \sum_{i=1}^{n} \big(\ (p_i \cdot A) + (p_i \cdot \Delta N_i)\ \big)$$

Separating Sum terms:

$$\overline{N} = \sum_{i=1}^{n} \ (p_i \cdot A) + \sum_{i=1}^{n} (p_i \cdot \Delta N_i)$$

$\boldsymbol{A}$ is a constant value which can be factored out:

$$\overline{N} = A \cdot \sum_{i=1}^{n} p_i + \sum_{i=1}^{n} (p_i \cdot \Delta N_i)$$

Again, we have constraint that: $\sum_{i=1}^{n} p_i = 1$

Therefore, the last expression becomes:

$$\overline{N} = A + \sum_{i=1}^{n} (p_i \cdot \Delta N_i)$$

This is the Weighted Differences method for Weighted Average and Expected Value calculations.

The variable $\boldsymbol{\Delta N_i}$ is a “Difference” from the **Anchor** (Reference value) and the expression $(\boldsymbol{p_1} \cdot \boldsymbol{\Delta N_i})$ is a **Weighted Difference** term**.** To calculate the Weighted Average we can sum the Weighted Differences and add them to the **Anchor** Value.

This will generally require $\boldsymbol{n}$ multiplications unless the Anchor value is selected as one of the values of $\boldsymbol{N}$. Note we are excluding cases where a Percentage Frequency is 0, as these can be ignored, so we have $\boldsymbol{n}$ values which have non-zero Percentage Frequencies or Weights.

Again, since:

$$\Delta N_i = N_i - A$$

Again, if the Anchor $\boldsymbol{A}$ is set equal to a particular value of $\boldsymbol{N}$ such as $\boldsymbol{N_j}$ for some $\boldsymbol{j}$, then $(\boldsymbol{\Delta N_j = 0})$:

$$\Delta N_j = N_j - A$$

By assumption:

$$A = N_j$$

Therefore

$$\Delta N_j = N_j - N_j = 0$$

Therefore, if the Anchor is set to any of the value of $\boldsymbol{N}$ such as $\boldsymbol{N_j}$ for some $\boldsymbol{j}$ then the Weighted Difference Method (Algorithm) will require $(\boldsymbol{n - 1})$ multiplication operations as the corresponding term in the Weighted Difference sum can be ignored as $(\boldsymbol{\Delta N_j = 0})$ and we skip that term in the sum and do not multiply by $\boldsymbol{p_j}$ as:

$p_j \cdot \Delta N_j = 0$ because $(\Delta N_j = 0)$

Proof for *FAST* Weighted Average Calculation Algorithm for Two Values

Here is a formal proof how why the *Fast* Weighted Average Algorithm for two values.

As a review, given two Values $(N_1,\ N_2)$ and corresponding Percentage Frequencies of $(p_1,\ p_2)$ (Normalized Weights), the traditional formula for Weighted Average of variable N $(\overline{N})$ when N can take with two Values is:

$$\overline{N} = \sum_{i=1}^{2} (p_i \cdot N_i) = (p_1 \cdot N_1) + (p_2 \cdot N_2)$$

$$\overline{N} = (N_1 \cdot p_1) + (N_2 \cdot p_2)$$

if either $\boldsymbol{p_1}$ or $\boldsymbol{p_2}$ is zero, then no multiplication is required as then either $\boldsymbol{p_1}$ or $\boldsymbol{p_2}$ is will be 1.

If they are not equal, without loss of generality, we can say $(\boldsymbol{N_2 \geq N_1})$ as we simply label the larger of two values $\boldsymbol{N_2}$. Define the variable $\boldsymbol{D}$ as Difference between $\boldsymbol{N_1}$ and $\boldsymbol{N_2}$:

$D = N_2 - N_1$ (D will be positive if $N_2 \geq N_1$)

Note that: $N_2 = N_1 + D$

Thus, by the definition, $\boldsymbol{D}$ will be positive. The Weighted Average of $(N_1,\ N_2)$, which is written as $(\overline{N})$ can be calculated as either:

$$\overline{N} = N_1 + (D \cdot p_2) \qquad \text{(Version 1)}$$

$$\overline{N} = N_1 - (D \cdot p_1) \qquad \text{(Version 2)}$$

We will start with our Traditional Weighted Average Formula, and transform it into both versions using the relationship $(p_1 + p_2 = 1)$.

$$\overline{N} = (N_1 \cdot p_1) + (N_2 \cdot p_2)$$

Inserting the expression for N_2 that $N_2 = N_1 + D$:

$$\overline{N} = (N_1 \cdot p_1) + (N_1 + D) \cdot p_2$$

Multiplying through by p_2:

$$\overline{N} = (N_1 \cdot p_1) + (N_1 \cdot p_2) + (D \cdot p_2)$$

Merging terms with N_1:

$$\overline{N} = N_1 \cdot (p_1 + p_2) + (D \cdot p_2)$$

Inserting expression $(p_1 + p_2 = 1)$:

$$\overline{N} = N_1 \cdot (1) + (D \cdot p_2)$$

$$\overline{N} = N_1 + (D \cdot p_2)$$

This is the expression we are looking for. We can also derive the other variant, using N_2 as the Reference value in a similar way. Note that: $(N_1 = N_2 - D)$:

$$\overline{N} = (N_1 \cdot p_1) + (N_2 \cdot p_2)$$

Inserting the expression for N_1:

$$\overline{N} = (N_2 - D) \cdot p_1 + (N_2 \cdot p_2)$$

Multiplying through by p_1 and moving D term to the end :

$$\overline{N} = (N_2 \cdot p_1) + (N_2 \cdot p_2) - (D \cdot p_1)$$

Merging terms with N_2:

$$\overline{N} = N_2 \cdot (p_1 + p_2) - (D \cdot p_1)$$

Inserting expression $(p_1 + p_2 = 1)$:

$$\overline{N} = N_2 \cdot (1) - (D \cdot p_1)$$

$$\overline{N} = N_2 - (D \cdot p_1)$$

Alternate Derivations

We can also derive these in a different way. Begin with:

$$\overline{N} = (N_1 \cdot p_1) + (N_2 \cdot p_2)$$

Insert expression $(p_1 = 1 - p_2)$ as $(p_1 + p_2 = 1)$

$$\overline{N} = N_1 \cdot (1 - p_2) + (N_2 \cdot p_2)$$

Multiply through by N_1:

$$\overline{N} = N_1 \cdot 1 - (N_1 \cdot p_2) + (N_2 \cdot p_2)$$

Rearrange terms:

$$\overline{N} = N_1 + (N_2 \cdot p_2) - (N_1 \cdot p_2)$$

Factor out a p_2:

$$\overline{N} = N_1 + (N_2 - N_1) \cdot p_2 = N_1 + (D \cdot p_2)$$

$$\overline{N} = N_1 + (p_2 \cdot D)$$

Similarly, begin with

$$\overline{N} = (N_1 \cdot p_1) + (N_2 \cdot p_2)$$

Insert expression $(p_2 = 1 - p_1)$ as $(p_1 + p_2 = 1)$

$$\overline{N} = (N_1 \cdot p_1) + N_2 \cdot (1 - p_1)$$

Multiply through by N_2:

$$\overline{N} = (N_1 \cdot p_1) + (N_2 \cdot 1) - (N_2 \cdot p_1)$$

Rearrange terms:

$$\overline{N} = N_2 + (N_1 \cdot p_1) - (N_2 \cdot p_1)$$

Factor out a p_1:

$$\overline{N} = N_2 + p_1 \cdot (N_1 - N_2)$$

Note that:

$$(N_1 - N_2) = -(N_2 - N_1) = -D$$

Therefore:

$$\overline{N} = N_2 + (p_2 \cdot (-D)) = N_2 - (p_2 \cdot D)$$

$$\overline{N} = N_2 - (p_1 \cdot D)$$

Proof of Difference-Based with MidPoint Formula (Average of Two Values)

Let's prove the Difference-based MidPoint formula starting with the Traditional Midpoint Formula:

Traditional Midpoint Formula: $$M = \frac{N_1 + N_2}{2}$$

The Fast MidPoint calculation algorithms we have are:

$$M = N_1 + \frac{D}{2} \qquad \text{[Version 1]}$$

$$M = N_2 - \frac{D}{2} \qquad \text{[Version 2]}$$

Where $(\boldsymbol{D = N_2 - N_1})$ and $(\boldsymbol{N_2 \geq N_1})$

Proof Method A

Version 1

Begin by inserting the expression for $\boldsymbol{D}$ in Version 1, which gives :

$$M = N_1 + \frac{N_2 - N_1}{2}$$

It should be noted that:

$$N_1 = \frac{2 \cdot N_1}{2}$$

Therefore:

$$M = N_1 + \frac{N_2 - N_1}{2} = \frac{2 \cdot N_1}{2} + \frac{N_2 - N_1}{2}$$

Combining terms onto a single fraction

$$M = \frac{2 \cdot N_1 + N_2 - N_1}{2}$$

Cancelling $\boldsymbol{N_1}$ terms:

$$M = \frac{N_1 + N_2}{2}$$

Version 2
We will also show the Version 2 is consistent with the Traditional MidPoint Formula. Version 2 is:

$$M = N_2 - \frac{D}{2} \qquad \text{where} \qquad D = (N_2 - N_1)$$

Inserting the expression for D into the equal above gives:

$$M = N_2 - \frac{N_2 - N_1}{2}$$

It should be noted that:

$$N_2 = \frac{2 \cdot N_2}{2} \quad \text{Therefore:}$$

$$M = N_2 - \frac{N_2 - N_1}{2} = \frac{2 \cdot N_2}{2} - \frac{N_2 - N_1}{2}$$

Combining terms onto a single fraction

$$M = \frac{2 \cdot N_2 - N_2 + N_1}{2}$$

Cancelling N_2 terms:

$$M = \frac{N_2 + N_1}{2}$$

Therefore these variations of the MidPoint Formula are mutually consistent.

Proof Method B

We can also go the other way and being with the Traditional MidPoint Formula and manipulate it:

$$M = \frac{N_1 + N_2}{2}$$

Version 1: Add and Subtract N_1 to the Numerator:

$$M = \frac{N_1 + N_2 + N_1 - N_1}{2}$$

Rearrange terms:

$$= \frac{(2 \cdot N_1) + (N_2 - N_1)}{2} = \frac{2 \cdot N_1}{2} + \frac{N_2 - N_1}{2}$$

$$M = N_1 + \frac{N_2 - N_1}{2} = N_1 + \frac{D}{2}$$

For Version 2, begin with:

$$M = \frac{N_1 + N_2}{2}$$

Then Add and Subtract $\boldsymbol{N_2}$ to the Numerator:

$$M = \frac{N_1 + N_2 + N_2 - N_2}{2} = \frac{N_1 + (2 \cdot N_2) - N_2}{2}$$

Rearrange terms:

$$M = \frac{N_1 - N_2 + (2 \cdot N_2)}{2} = \frac{N_1 - N_2}{2} + \frac{(2 \cdot N_2)}{2} = \frac{N_1 - N_2}{2} + N_2$$

$$M = N_2 + \frac{N_1 - N_2}{2} = N_2 - \frac{N_2 - N_1}{2}$$

$$M = N_2 - \frac{D}{2}$$

Proof Method C

The Fast MidPoint Formula can also be calculated from the Fast Weighted Average Formula for two values which were:

$$\overline{N} = N_1 + (D \cdot p_2) \qquad \text{(Version 1)}$$

$$\overline{N} = N_1 - (D \cdot p_1) \qquad \text{(Version 2)}$$

Where $(\boldsymbol{D = N_2 - N_1})$ and $(\boldsymbol{N_2 \geq N_1})$

Note that in the MIdPoint or Average of two values, this is equivalent to the Weighted Average Formula where each weight is $50\,\%$:

$$M = \frac{N_1 + N_2}{2} = \frac{N_1}{2} + \frac{N_2}{2}$$

$$M = \left(\tfrac{1}{2} \cdot N_1\right) + \left(\tfrac{1}{2} \cdot N_2\right)$$

Since $\frac{1}{2} = 50\,\%$:

$$M = (50\% \cdot N_1) + (50\% \cdot N_2)$$

This means the MidPoint is the Weighted Average of two values given equal weights so $(p_1 = p_2 = 50\%)$. The MidPoint is therefore :

$$M = N_1 + (50\% \cdot D) = N_1 + d = N_1 \ + \ \tfrac{D}{2}$$

$$M = N_2 - (50\% \cdot D) = N_2 - d = N_2 \ - \ \tfrac{D}{2}$$

FAST Direct Average Algorithm

The Direct Average of $\boldsymbol{n}$ data points or Values of variable $\boldsymbol{x}$ with values $\{x_1, x_2, \ \ldots \ x_i, \ \ldots \ x_n\}$ is denoted by $\boldsymbol{\bar{x}}$ and calculated as:

$$\text{Average of } x\text{:} \quad \bar{x} = \frac{\sum_{i=1}^{n}(x_i)}{n}$$

The Direct Average of a set of numbers is the Unweighted Average of the set of numbers, and can also be thought of as the Weighted Average of the set of Values where each value is given **Equal Weight:**

$$\bar{x} = \frac{\sum_{i=1}^{n}(x_i)}{n} = \sum_{i=1}^{n}\left(\frac{1}{n}\cdot x_i\right) \qquad \text{[The Weight of each value is } \tfrac{1}{n}\text{]}$$

These weights are normalized as:

$$= \sum_{i=1}^{n}\left(\frac{1}{n}\right) = n\cdot\frac{1}{n} = 1$$

Therefore, we can apply our *Fast* Weighted Average Formula with weight $\frac{1}{n}$. We can select any value as a Reference Value or an Arbitrary Reference Value. Selecting a Value as the Reference Value and labeling it $\boldsymbol{x_1}$ gives:

$$\bar{x} = x_1 + \frac{\sum_{i=2}^{n}(x_i - x_1)}{n}$$

Selecting an arbitrary Reference Value $\boldsymbol{A}$ gives:

$$\bar{x} = A + \frac{\sum_{i=1}^{n}(x_i - A)}{n}$$

The version using $\boldsymbol{x_1}$ as the Reference Value makes the term $(x_1 - A)$ equal to zero, and so this is excluded from the sum.

Proof

Again, set the reference value equal to $\boldsymbol{A}$, and express each value relative to $\boldsymbol{A}$:

$$\Delta x_i = x_i - \boldsymbol{A} \qquad \Longrightarrow \qquad x_i = \boldsymbol{A} + \Delta x_i$$

$$\bar{x} = \frac{\sum_{i=1}^{n}(x_i)}{n} = \frac{\sum_{i=1}^{n}(A + \Delta x_i)}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} A}{n} + = \frac{\sum_{i=1}^{n} + \Delta x_i}{n}$$

$$\bar{x} = \frac{n \cdot A}{n} + \frac{\sum_{i=1}^{n} \Delta x_i}{n}$$

$$\bar{x} = A + \frac{\sum_{i=1}^{n} \Delta x_i}{n} = A + \frac{\sum_{i=1}^{n} (x_i - A)}{n}$$

Setting A equal to x_1 eliminates the first term in the sum, so doing so allows the sum to run from $\boldsymbol{i}$ equals 2 to $\boldsymbol{n}$ as the first term is zero.

Assumptions

Throughout this paper, I have assumed that none of the Normalized Weights $\boldsymbol{w_i}$ or Percentage Frequencies $\boldsymbol{p_i}$ values are 0 or negative. If any are zero, then they will ignored and excluded from the count of $\boldsymbol{n}$. So we will assume $\boldsymbol{n}$ positive weights $\{w_1, w_2, \ \ldots \ w_i, \ \ldots \ w_n\}$. So the number $\boldsymbol{n}$ refers to the number of Values that remain after discarding values with zero Weight.

Arbitrary Weights

A similar method can be used if arbitrary (non-formalized) weights are given to calculate the Weighted average in at most $(\boldsymbol{n-1})$ multiplication operations given $\boldsymbol{n}$ positive arbitrary Weights $\{W_1, W_2, \ \ldots \ W_i, \ \ldots \ W_n\}$ and $\boldsymbol{n}$ Values of variable $\boldsymbol{x}$ with Values $\{x_1, x_2, \ \ldots \ x_i, \ \ldots \ x_n\}$. Again, the Weights must be non-negative, and any Weights which are equal to zero will be discarded long with the corresponding Value.

The Weighted Average of variable $\boldsymbol{x}$ given arbitrary (non-Normalized) Weights is defined as:

$$\bar{x} = \frac{\sum_{i=1}^{n} (W_i \cdot x_i)}{\sum_{i=1}^{n} W_i}$$

Defining the variable $\boldsymbol{W_S}$ ("S" is for "Sum") as:

$$W_S = \sum_{i=1}^{n} W_i$$

We can rewrite the equation for the Weighted Average as:

$$\bar{x} = \frac{\sum_{i=1}^{n} (W_i \cdot x_i)}{W_S}$$

In general, the formula above will require $\boldsymbol{n}$ multiplication operations and a division operation at the end to divide by $\boldsymbol{W_S}$. Normalizing the Weights will also require $\boldsymbol{n}$ division operations to divide each original Weight by the sum of the Weights $(\boldsymbol{W_S})$.

We can calculate the Weighted Average without Normalizing the weights and with at most $(\boldsymbol{n-1})$ multiplication operations through the formula below. Again, assume that we use the value $\boldsymbol{x_1}$ as the Reference Value.

Again, define $\boldsymbol{\Delta x_i}$ as follows:

$$\Delta x_i = x_i - \boldsymbol{x_1} \qquad \Longleftrightarrow \qquad x_i = \boldsymbol{x_1} + \Delta x_i$$

The Weighted Average of $\boldsymbol{x}$ can be calculated as:

$$\bar{x} = \frac{\sum_{i=2}^{n} (W_i \cdot \Delta x_i)}{W_S} + x_1$$

Note the sum runs from $\boldsymbol{i}$ going from 2 to $\boldsymbol{n}$, and $\boldsymbol{W_S}$ is outside of the sum, so we can calculating the sum of the Weighted Differences and divide once by $\boldsymbol{W_S}$. Therefore, implementing this formula will require at most $(\boldsymbol{n-1})$ multiplication operations and a single Division calculation.

Pseudo-Code

Pseudo-code for implementing this calculation algorithm is shown below.

Assume the inputs are arrays W[n] and x[n] which contain floating point numbers for the Weights and Values respectively. We will also assume x[0] is the Reference Value. Each array is zero-indexed and has **n** values. We also use the arrays D[n] for the Difference values, and WD[n] for the Weighted Difference values.

```
WD_Sum=0

W_Sum= W[0]

for i = 1 to (n-1) {

    D[i]=x[i]-x[0];

    WD[i]=W[i]*x[i];

    WD_Sum = WD_Sum+WD[i]

    W_Sum=W_Sum+W[i];

}

WA=x[0]+ (WD_Sum / W_Sum);
```



```
Return WA;
```

The above calculations can also be done in-place to not have additional arrays D[n] and WD[n]. That is when retaining the Weights, if you have the Weighted Difference and the Reference Value then one can reconstruct the original set of values given. Pseudo-code for in-place calculations is given below.

```
WD_Sum=0

W_Sum= W[0]

for i = 1 to (n-1) {

    x[i]=x[i]-x[0];

    x[i]=W[i]*x[i];

    WD_Sum = WD_Sum+x[i]

    W_Sum=W_Sum+W[i];

}

WA=x[0]+ (WD_Sum / W_Sum);

Return WA;
```

For both of these, the code can be modified if the Weights are Normalized to simply not calculate the W_Sum value, and not divide by the W_Sum value, since this will always be 1 given the Weights are normalized. Pseudocode for normalized weights is shown below:

```
WD_Sum=0

for i = 1 to (n-1) {

    x[i]=x[i]-x[0];

    x[i]=w[i]*x[i];

    WD_Sum = WD_Sum+x[i]

}

WA=x[0]+ WD_Sum

Return WA;
```


Arbitrary Reference Value

We can also use an Arbitrary Reference Value $\boldsymbol{A}$. Again, in general, an Arbitrary Reference Value which is not one of the values of $\boldsymbol{x_i}$ will require will require $\boldsymbol{n}$ multiplication operations.

Define the Reference Value as $\boldsymbol{A}$ ("**A**" is for **A**nchor), which can be **any value**.

Define $\boldsymbol{\Delta x_i}$ as follows:

$$\Delta x_i = x_i - \boldsymbol{A} \qquad \Longleftrightarrow \qquad x_i = \Delta x_I + A$$

The Weighted Average of $\boldsymbol{x}$ can be calculated as:

$$\bar{x} = \frac{\sum_{i=1}^{n} (W_i \cdot \Delta x_i)}{W_S} + A$$

Again, setting $\boldsymbol{A}$ equals to $\boldsymbol{x_j}$ will make $(\boldsymbol{\Delta x_j = 0})$ so we can ignore the corresponding term in the sum as we are multiplying by zero.

Pseudo-code for implementing this calculation algorithm is shown below. Inputs are arrays W[n] and x[n] which contain floating point numbers for the Weights, and Values respectively, and a separate variable A for the Reference Value.

```
WD_Sum=0

W_Sum= W[0]

for i = 0 to (n-1) {

        x[i]=x[i]-A;

        x[i]=W[i]*x[i];

        WD_Sum = WD_Sum+x[i]

        W_Sum=W_Sum+W[i];

}

WA=A + (WD_Sum / W_Sum);

Return WA;
```

Note that in this example, the sum runs from i going from 0 to n-1. Again, the are zero-indexed as per computer science standards.

Pseudocde for arbitrary Reference Values given Normalized Weights is shown below:

```
WD_Sum=0

for i = 0 to (n-1) {

    x[i]=x[i]-A;

    x[i]=w[i]*x[i];

    WD_Sum = WD_Sum+x[i]

}

WA = A + WD_Sum;

Return WA;
```

This can also be further refined to save an operation and an intermediate variable as:

```
Temp = A

for i = 0 to (n-1) {

    x[i]=x[i]-A;

    x[i]=w[i]*x[i];

    Temp = Temp+x[i]

}

Return Temp;
```

