# Incremental calculation of weighted mean and variance


Tony Finch

⟨fanf2@cam.ac.uk⟩ ⟨dot@dotat.at⟩

University of Cambridge Computing Service


February 2009


**Abstract**

In these notes I explain how to derive formulae for numerically stable calculation of the mean and standard deviation, which are also suitable for incremental on-line calculation. I then generalize these formulae to weighted means and standard deviations. I unpick the difficulties that arise when generalizing further to normalized weights. Finally I show that the exponentially weighted moving average is a special case of the incremental normalized weighted mean formula, and derive a formula for the exponentially weighted moving standard deviation.


## 1 Simple mean

Straightforward translation of equation 1 into code can suffer from loss of precision because of the difference in magnitude between a sample and the sum of all samples. Equation 4 calculates the mean in a way that is more numerically stable because it avoids accumulating large sums.

$$
\begin{aligned}
\mu_n &= \frac{1}{n}\sum_{i=1}^{n} x_i && (1)\\
&= \frac{1}{n}(x_n + \sum_{i=1}^{n-1} x_i) && (2)\\
&= \frac{1}{n}(x_n + (n-1)\mu_{n-1}) && (3)\\
&= \mu_{n-1} + \frac{1}{n}(x_n - \mu_{n-1}) && (4)
\end{aligned}
$$

This formula also provides us with some useful identities.

$$
\begin{aligned}
x_n - \mu_{n-1} &= n(\mu_n - \mu_{n-1}) && (5)\\
x_n - \mu_n &= n(\mu_n - \mu_{n-1}) - \mu_n + \mu_{n-1} &&\\
&= (n-1)(\mu_n - \mu_{n-1}) && (6)
\end{aligned}
$$

## 2 Simple variance

The definition of the standard deviation in equation 7 below requires us to already know the mean, which implies two passes over the data. This isn't feasible for online algorithms that need to produce incremental results after each sample becomes available. Equation 12 solves this problem since it allows us to calculate the standard deviation from two running sums.

$$
\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 && (7)\\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\mu + \mu^2) && (8)
\end{aligned}
$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\mu\frac{1}{n}\sum_{i=1}^{n} x_i + \mu^2\frac{1}{n}\sum_{i=1}^{n} 1 \tag{9}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\mu\mu + \mu^2\frac{n}{n} \tag{10}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^2 \tag{11}$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 \tag{12}$$

# 3 Incremental variance

Knuth notes [1] that equation 12 is prone to loss of precision because it takes the difference between two large sums of similar size, and suggests equation 24 as an alternative that avoids this problem. However he does not say how it is derived. In the following, equation 20 is derived from the previous step using equation 5.

$$
\begin{aligned}
\text{Let } S_n &= n\sigma_n^2 && (13)\\
&= \textstyle\sum_{i=1}^{n}(x_i - \mu_n)^2 && (14)\\
&= \textstyle\sum_{i=1}^{n} x_i^2 - n\mu_n^2 && (15)\\
S_n - S_{n-1} &= \sum_{i=1}^{n} x_i^2 - n\mu_n^2 - \sum_{i=1}^{n-1} x_i^2 + (n-1)\mu_{n-1}^2 && (16)\\
&= x_n^2 - n\mu_n^2 + (n-1)\mu_{n-1}^2 && (17)\\
&= x_n^2 - \mu_{n-1}^2 + n(\mu_{n-1}^2 - \mu_n^2) && (18)\\
&= x_n^2 - \mu_{n-1}^2 + n(\mu_{n-1} - \mu_n)(\mu_{n-1} + \mu_n) && (19)\\
&= x_n^2 - \mu_{n-1}^2 + (\mu_{n-1} - x_n)(\mu_{n-1} + \mu_n) && (20)\\
&= x_n^2 - \mu_{n-1}^2 + \mu_{n-1}^2 - x_n\mu_n - x_n\mu_{n-1} + \mu_n\mu_{n-1} && (21)\\
&= x_n^2 - x_n\mu_n - x_n\mu_{n-1} + \mu_n\mu_{n-1} && (22)\\
&= (x_n - \mu_{n-1})(x_n - \mu_n) && (23)\\
S_n &= S_{n-1} + (x_n - \mu_{n-1})(x_n - \mu_n) && (24)\\
\sigma_n &= \sqrt{S_n/n} && (25)
\end{aligned}
$$

Mathworld [2] has an alternative derivation of a similar formula, which in our notation is as follows.

$$
\begin{aligned}
S_n &= \sum_{i=1}^{n}(x_i - \mu_n)^2 && (26)\\
&= \sum_{i=1}^{n}((x_i - \mu_{n-1}) - (\mu_n - \mu_{n-1}))^2 && (27)\\
&= \sum_{i=1}^{n}(x_i - \mu_{n-1})^2 + \sum_{i=1}^{n}(\mu_n - \mu_{n-1})^2 - 2\sum_{i=1}^{n}(x_i - \mu_{n-1})(\mu_n - \mu_{n-1}) && (28)
\end{aligned}
$$

Simplify the first summation:

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \mu_{n-1})^2 &= (x_n - \mu_{n-1})^2 + \sum_{i=1}^{n-1}(x_i - \mu_{n-1})^2 && (29)\\
&= (x_n - \mu_{n-1})^2 + S_{n-1} && (30)\\
&= S_{n-1} + n^2(\mu_n - \mu_{n-1})^2 && (31)
\end{aligned}
$$

Simplify the second summation:

$$\sum_{i=1}^{n}(\mu_n - \mu_{n-1})^2 \quad = \quad n(\mu_n - \mu_{n-1})^2 \tag{32}$$

Simplify the third summation:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \mu_{n-1})(\mu_n - \mu_{n-1}) &= (\mu_n - \mu_{n-1})\sum_{i=1}^{n}(x_i - \mu_{n-1}) && (33)\\
&= (\mu_n - \mu_{n-1})\left(x_n - \mu_{n-1} + \sum_{i=1}^{n-1}(x_i - \mu_{n-1})\right) && (34)\\
&= (\mu_n - \mu_{n-1})\left(x_n - \mu_{n-1} - (n-1)\mu_{n-1} + \sum_{i=1}^{n-1} x_i\right) && (35)\\
&= (\mu_n - \mu_{n-1})(x_n - \mu_{n-1} - (n-1)\mu_{n-1} + (n-1)\mu_{n-1}) && (36)\\
&= (\mu_n - \mu_{n-1})(x_n - \mu_{n-1}) && (37)\\
&= n(\mu_n - \mu_{n-1})^2 && (38)
\end{aligned}$$

Back to the complete formula:

$$\begin{aligned}
S_n &= S_{n-1} + n^2(\mu_n - \mu_{n-1})^2 + n(\mu_n - \mu_{n-1})^2 - 2n(\mu_n - \mu_{n-1})^2 && (39)\\
&= S_{n-1} + n^2(\mu_n - \mu_{n-1})^2 - n(\mu_n - \mu_{n-1})^2 && (40)\\
&= S_{n-1} + n(n-1)(\mu_n - \mu_{n-1})^2 && (41)
\end{aligned}$$

We can use equations 6 and 5 to show this is equivalent to equation 24.

$$\begin{aligned}
S_n &= S_{n-1} + n(n-1)(\mu_n - \mu_{n-1})^2 && (42)\\
&= S_{n-1} + n(\mu_n - \mu_{n-1})(x_n - \mu_n) && (43)\\
&= S_{n-1} + (x_n - \mu_{n-1})(x_n - \mu_n) && (44)
\end{aligned}$$

# 4 Weighted mean

The weighted mean is defined as follows.

$$\mu = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{45}$$

It is equivalent to the simple mean when all the weights are equal, since

$$\mu = \frac{\sum_{i=1}^{n} w x_i}{\sum_{i=1}^{n} w} = \frac{w\sum_{i=1}^{n} x_i}{nw} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{46}$$

If the samples are all different, then weights can be thought of as sample frequencies, or they can be used to calculate probabilities where $p_i = w_i / \sum w_i$. The following derivation of the incremental formula (equation 53) follows the same pattern as the derivation of equation 4. For brevity we also define $W_n$ as the sum of the weights.

$$\begin{aligned}
W_n &= \sum_{i=1}^{n} w_i && (47)\\
\mu_n &= \frac{1}{W_n}\sum_{i=1}^{n} w_i x_i && (48)\\
&= \frac{1}{W_n}\left(w_n x_n + \sum_{i=1}^{n-1} w_i x_i\right) && (49)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{W_n}(w_n x_n + W_{n-1}\mu_{n-1}) && (50)\\
&= \frac{1}{W_n}(w_n x_n + (W_n - w_n)\mu_{n-1}) && (51)\\
&= \frac{1}{W_n}(W_n\mu_{n-1} + w_n x_n - w_n\mu_{n-1}) && (52)\\
&= \mu_{n-1} + \frac{w_n}{W_n}(x_n - \mu_{n-1}) && (53)
\end{aligned}
$$

Useful identities derived from this formula are:

$$
\begin{aligned}
W_n(\mu_{n-1} - \mu_n) &= w_n(\mu_{n-1} - x_n) && (54)\\
\frac{W_n}{w_n}(\mu_n - \mu_{n-1}) &= x_n - \mu_{n-1} && (55)\\
x_n - \mu_n &= \frac{W_n}{w_n}(\mu_n - \mu_{n-1}) - \mu_n + \mu_{n-1} \\
&= \frac{W_n - w_n}{w_n}(\mu_n - \mu_{n-1}) && (56)\\
&= \frac{W_n - w_n}{W_n}(x_n - \mu_{n-1}) && (57)
\end{aligned}
$$

# 5 Weighted variance

Similarly, we derive a numerically stable formula for calculating the weighted variance (equation 68) using the same pattern as the derivation of the unweighed wersion (equation 24).

$$
\begin{aligned}
\sigma^2 &= \frac{1}{W_n}\sum_{i=1}^{n} w_i(x_i - \mu)^2 = \frac{1}{W_n}\sum_{i=1}^{n} w_i x_i^2 - \mu^2 && (58)\\
\text{Let } S_n &= W_n\sigma_n^2 && (59)\\
&= \sum_{i=1}^{n} w_i x_i^2 - W_n\mu_n^2 && (60)\\
S_n - S_{n-1} &= \sum_{i=1}^{n} w_i x_i^2 - W_n\mu_n^2 - \sum_{i=1}^{n-1} w_i x_i^2 + W_{n-1}\mu_{n-1}^2 && (61)\\
&= w_n x_n^2 - W_n\mu_n^2 + W_{n-1}\mu_{n-1}^2 && (62)\\
&= w_n x_n^2 - W_n\mu_n^2 + (W_n - w_n)\mu_{n-1}^2 && (63)\\
&= w_n\left(x_n^2 - \mu_{n-1}^2\right) + W_n(\mu_{n-1}^2 - \mu_n^2) && (64)\\
&= w_n\left(x_n^2 - \mu_{n-1}^2\right) + W_n(\mu_{n-1} - \mu_n)(\mu_{n-1} + \mu_n) && (65)\\
&= w_n\left(x_n^2 - \mu_{n-1}^2 + (\mu_{n-1} - x_n)(\mu_{n-1} + \mu_n)\right) && (66)\\
&= w_n(x_n - \mu_{n-1})(x_n - \mu_n) && (67)\\
S_n &= S_{n-1} + w_n(x_n - \mu_{n-1})(x_n - \mu_n) && (68)\\
\sigma_n &= \sqrt{S_n/W_n} && (69)
\end{aligned}
$$

# 6 Variable weights

In the previous three sections, I have assumed that weights are constant once assigned. However a common requirement is to normalize the weights, such that

$$
W = \sum_{i=1}^{n} w_i = 1 \tag{70}
$$

If we are repeatedly adding new data to our working set, then we can't have both constant weights and normalized weights. To allow us to keep weights normalized, we need to allow the weight of each sample to change as the set of samples changes. To indicate this we will give weights two indices, the first identifying the set of samples using the sample count (as we have been doing for $\mu_n$ etc.) and the second being the index of the sample in that set. We will not make any assumptions about the sum of the weights, that is we will not require them to be normalized. For example,

$$W_n = \sum_{i=1}^{n} w_{n,i}, \;\; \mu_n = \frac{1}{W_n} \sum_{i=1}^{n} w_{n,i} x_i \tag{71}$$

Having done this we need to re-examine some of the logical steps in the previous sections to ensure they are still valid. In equations 49–51, we used the fact that in the fixed-weight setting,

$$\sum_{i=1}^{n-1} w_i x_i = W_{n-1}\mu_{n-1} = (W_n - w_n)\mu_{n-1} \tag{72}$$

In the new setting, this equality is fairly obviously no longer true. (For example, if we are keeping weights normalized then $W_n = W_{n-1} = 1$.) Fortunately there is a different middle step which justifies equation 72 when weights vary, so the results of section 4 remain valid.

$$\begin{aligned}
\sum_{i=1}^{n-1} w_{n,i} x_i &= (W_n - w_{n,n})\mu_{n-1} && (73)\\
\sum_{i=1}^{n-1} w_{n,i} x_i &= \sum_{i=1}^{n-1} w_{n,i} \frac{\sum_{i=1}^{n-1} w_{n-1,i} x_i}{\sum_{i=1}^{n-1} w_{n-1,i}} && (74)\\
\frac{\sum_{i=1}^{n-1} w_{n,i} x_i}{\sum_{i=1}^{n-1} w_{n,i}} &= \frac{\sum_{i=1}^{n-1} w_{n-1,i} x_i}{\sum_{i=1}^{n-1} w_{n-1,i}} && (75)\\
\sum_{i=1}^{n-1} \frac{w_{n,i}}{\sum_{i=1}^{n-1} w_{n,i}} x_i &= \sum_{i=1}^{n-1} \frac{w_{n-1,i}}{\sum_{i=1}^{n-1} w_{n-1,i}} x_i && (76)\\
\frac{w_{n,j}}{\sum_{i=1}^{n-1} w_{n,i}} &= \frac{w_{n-1,j}}{\sum_{i=1}^{n-1} w_{n-1,i}} \text{ where } 1 \le j \le n-1 && (77)
\end{aligned}$$

This says that for the weighted mean formulae to remain valid the new and old weights should be consistent. Equation 75 says that we get the same result when we calculate the mean of the previous working set whether we use the old weights or the new weights. Equation 77 says that when we normalize the weights across the previous set (up to $n-1$) we get the same set of weights whether we start from the old weights or the new ones. This requirement isn't enough by itself to make the weighted variance formulae work, so we will examine it again below.

# 7 The expectation function

At this point it is worth defining some better notation to reduce the number of summations we need to write. The expectation function is a generalized version of the mean, whose argument is some arbitrary function of each sample.

$$\begin{aligned}
E_n(f(x)) &= \frac{1}{W_n} \sum_{i=1}^{n} w_{n,i} f(x_i) && (78)\\
E_n(k) &= k && (79)\\
E_n(af(x)) &= aE_n(f(x)) && (80)\\
E_n(f(x)+g(x)) &= E_n(f(x)) + E_n(g(x)) && (81)\\
\mu_n &= E_n(x) && (82)\\
\sigma_n^2 &= E_n((x-\mu_n)^2) && (83)\\
&= E_n(x^2 + \mu_n^2 - 2\mu_n x) && (84)\\
&= E_n(x^2) + \mu_n^2 - 2\mu_n E_n(x) && (85)
\end{aligned}$$

$$
\begin{aligned}
&= E_n(x^2) - \mu_n^2 && (86)\\
&= E_n(x^2) - E_n(x)^2 && (87)
\end{aligned}
$$

The incremental formula is derived in the usual way. Equation 92 is particularly useful.

$$
\begin{aligned}
W_n E_n(f(x)) &= \sum_{i=1}^{n} w_{n,i} f(x_i) && (88)\\
&= w_{n,n} f(x_n) + \sum_{i=1}^{n-1} w_{n,i} f(x_i) && (89)\\
&= w_{n,n} f(x_n) + (W_n - w_{n,n}) \frac{\sum_{i=1}^{n-1} w_{n,i} f(x_i)}{\sum_{i=1}^{n-1} w_{n,i}} && (90)\\
&= w_{n,n} f(x_n) + (W_n - w_{n,n}) \frac{\sum_{i=1}^{n-1} w_{n-1,i} f(x_i)}{\sum_{i=1}^{n-1} w_{n-1,i}} && (91)\\
&= w_{n,n} f(x_n) + (W_n - w_{n,n}) E_{n-1}(f(x)) && (92)\\
E_n(f(x)) &= E_{n-1}(f(x)) + \frac{w_{n,n}}{W_n} \left(f(x_n) - E_{n-1}(f(x))\right) && (93)
\end{aligned}
$$

# 8 Variable-weight variance

In equations 61–63 we made the following assumptions which are not true when weights can vary.

$$
\begin{aligned}
&\sum_{i=1}^{n} w_{n,i} x_i^2 - W_n \mu_n^2 - \sum_{i=1}^{n-1} w_{n-1,i} x_i^2 + W_{n-1} \mu_{n-1}^2\\
\neq\ & w_{n,n} x_n^2 - W_n \mu_n^2 + W_{n-1} \mu_{n-1}^2\\
\neq\ & w_{n,n} x_n^2 - W_n \mu_n^2 + (W_n - w_{n,n}) \mu_{n-1}^2
\end{aligned}
$$

If we try to re-do the short derivation of the incremental standard deviation formula starting from $S_n - S_{n-1}$ then we soon get stuck. Fortunately the longer derivation shows how to made it work.

$$
\begin{aligned}
S_n &= W_n \sigma_n^2 && (94)\\
&= W_n E_n\left((x - \mu_n)^2\right) && (95)\\
&= W_n E_n\left(([x - \mu_{n-1}] - [\mu_n - \mu_{n-1}])^2\right) && (96)\\
&= W_n E_n\left([x - \mu_{n-1}]^2 + [\mu_n - \mu_{n-1}]^2 - 2[x - \mu_{n-1}][\mu_n - \mu_{n-1}]\right) && (97)\\
&= W_n E_n\left([x - \mu_{n-1}]^2\right) + W_n E_n\left([\mu_n - \mu_{n-1}]^2\right) - 2 W_n E_n\left([x - \mu_{n-1}][\mu_n - \mu_{n-1}]\right) && (98)
\end{aligned}
$$

Simplify the first term:

$$
\begin{aligned}
W_n E_n\left([x - \mu_{n-1}]^2\right) &= w_{n,n}[x_n - \mu_{n-1}]^2 + (W_n - w_{n,n}) E_{n-1}\left([x - \mu_{n-1}]^2\right) && (99)\\
&= w_{n,n}[x_n - \mu_{n-1}]^2 + (W_n - w_{n,n}) \frac{S_{n-1}}{W_{n-1}} && (100)\\
&= \frac{W_n - w_{n,n}}{W_{n-1}} S_{n-1} + \frac{W_n^2}{w_{n,n}} [\mu_n - \mu_{n-1}]^2 && (101)
\end{aligned}
$$

Simplify the second term:

$$
W_n E_n\left([\mu_n - \mu_{n-1}]^2\right) = W_n [\mu_n - \mu_{n-1}]^2 \qquad (102)
$$

Simplify the third term:

$$
\begin{aligned}
& W_n E_n([x - \mu_{n-1}][\mu_n - \mu_{n-1}])\\
=\ & [\mu_n - \mu_{n-1}] W_n E_n [x - \mu_{n-1}] && (103)
\end{aligned}
$$

$$
\begin{aligned}
&= [\mu_n - \mu_{n-1}]\left(w_{n,n}[x_n - \mu_{n-1}] + (W_n - w_{n,n})E_{n-1}[x - \mu_{n-1}]\right) && (104)\\
&= [\mu_n - \mu_{n-1}]\left(w_{n,n}[x_n - \mu_{n-1}] + (W_n - w_{n,n})[E_{n-1}(x) - E_{n-1}(\mu_{n-1})]\right) && (105)\\
&= [\mu_n - \mu_{n-1}]\left(w_{n,n}[x_n - \mu_{n-1}] + (W_n - w_{n,n})[\mu_{n-1} - \mu_{n-1}]\right) && (106)\\
&= [\mu_n - \mu_{n-1}]w_{n,n}[x_n - \mu_{n-1}] && (107)\\
&= W_n[\mu_n - \mu_{n-1}]^2 && (108)
\end{aligned}
$$

Back to the complete formula:

$$
\begin{aligned}
S_n &= \frac{W_n - w_{n,n}}{W_{n-1}}S_{n-1} + \frac{W_n^2}{w_{n,n}}[\mu_n - \mu_{n-1}]^2 + W_n[\mu_n - \mu_{n-1}]^2 - 2W_n[\mu_n - \mu_{n-1}]^2 && (109)\\
&= \frac{W_n - w_{n,n}}{W_{n-1}}S_{n-1} + \frac{W_n^2}{w_{n,n}}[\mu_n - \mu_{n-1}]^2 - \frac{W_n w_{n,n}}{w_{n,n}}[\mu_n - \mu_{n-1}]^2 && (110)\\
&= \frac{W_n - w_{n,n}}{W_{n-1}}S_{n-1} + (W_n - w_{n,n})\frac{W_n}{w_{n,n}}[\mu_n - \mu_{n-1}]^2 && (111)\\
&= \frac{W_n - w_{n,n}}{W_{n-1}}S_{n-1} + (W_n - w_{n,n})[\mu_n - \mu_{n-1}](x_n - \mu_{n-1}) && (112)\\
S_n &= \frac{W_n - w_{n,n}}{W_{n-1}}S_{n-1} + w_{n,n}(x_n - \mu_n)(x_n - \mu_{n-1}) && (113)
\end{aligned}
$$

This is the same as equation 68, except for the multiplier $\frac{W_n - w_{n,n}}{W_{n-1}}$ which captures the change in weights between the old and new sets.

$$
\frac{W_n - w_{n,n}}{W_{n-1}} = \frac{\sum_{i=1}^{n-1} w_{n,i}}{\sum_{i=1}^{n-1} w_{n-1,i}} = \frac{w_{n,j}}{w_{n-1,j}} \text{ where } 1 \le j \le n-1 \qquad (114)
$$

Now that we know the rescaling trick which makes it work, we can write down the short version.

$$
\begin{aligned}
& S_n - \frac{W_n - w_{n,n}}{W_{n-1}}S_{n-1} \\
&= W_n\left(E_n(x^2) - \mu_n^2\right) - (W_n - w_{n,n})\left(E_{n-1}(x^2) - \mu_{n-1}^2\right) && (115)\\
&= W_n\left(E_n(x^2) - \mu_n^2\right) - W_n E_n(x^2) + w_{n,n}x_n^2 + (W_n - w_{n,n})\mu_{n-1}^2 && (116)\\
&= w_{n,n}x_n^2 - W_n\mu_n^2 + (W_n - w_{n,n})\mu_{n-1}^2 && (117)\\
&= w_{n,n}\left(x_n^2 - \mu_{n-1}^2\right) + W_n(\mu_{n-1}^2 - \mu_n^2) && (118)\\
&= w_{n,n}\left(x_n^2 - \mu_{n-1}^2\right) + W_n(\mu_{n-1} - \mu_n)(\mu_{n-1} + \mu_n) && (119)\\
&= w_{n,n}\left(x_n^2 - \mu_{n-1}^2 + (\mu_{n-1} - x_n)(\mu_{n-1} + \mu_n)\right) && (120)\\
&= w_{n,n}(x_n - \mu_n)(x_n - \mu_{n-1}) && (121)
\end{aligned}
$$

# 9 Exponentially-weighted mean and variance

Starting from equation 53, let's set $w_{n,n}/W_n$ to a constant $0 < \alpha < 1$ and let $a = 1 - \alpha$. This produces the standard formula for the exponentially weighted moving average.

$$
\begin{aligned}
\mu_n &= \mu_{n-1} + \alpha(x_n - \mu_{n-1}) && (122)\\
&= (1 - \alpha)\mu_{n-1} + \alpha x_n && (123)\\
&= a\mu_{n-1} + (1 - a)x_n && (124)
\end{aligned}
$$

In the following it's more convenient to use a lower bound of 0 instead of 1, i.e. $0 \le i \le n$. We are going to show that the weights are renormalized each time a datum is added. First, we expand out the inductive definition of the mean.

$$
\begin{aligned}
\mu_n &= a\mu_{n-1} + (1 - a)x_n && (125)\\
&= a^2\mu_{n-2} + a(1 - a)x_{n-1} + (1 - a)x_n && (126)
\end{aligned}
$$

$$
\begin{aligned}
&= a^3\mu_{n-3} + a^2(1-a)x_{n-2} + a(1-a)x_{n-1} + (1-a)x_n && (127)\\
\mu_n &= a^n x_0 + \sum_{i=1}^{n} a^{n-i}(1-a)x_i && (128)
\end{aligned}
$$

This allows us to write down the weights directly. Note that $w_{n,n}$ is independent of $n$.

$$
\begin{aligned}
w_{n,0} &= a^n && (129)\\
w_{n,i} &= a^{n-i}(1-a), \text{ for } 1 \le i \le n && (130)\\
w_{n,n} &= 1-a \;=\; \alpha && (131)
\end{aligned}
$$

Since $w_{n,n} = \alpha = w_{n,n}/W_n$ we can see that $\forall n.\ W_n = 1$, that is, the weights are always normalized. We can get the same result by summing the geometric series.

$$
\sum_{i=1}^{n} a^{n-i} = \sum_{j=0}^{n-1} a^j = \frac{1-a^n}{1-a} \tag{132}
$$

$$
\sum_{i=1}^{n} w_{n,i} = \sum_{i=1}^{n} a^{n-i}(1-a) = 1-a^n \tag{133}
$$

$$
W_n = w_{n,0} + \sum_{i=1}^{n} w_{n,i} = a^n + (1-a^n) = 1 \tag{134}
$$

These weights satisfy the consistency requirement because

$$
\frac{w_{n,j}}{\sum_{i=1}^{n-1} w_{n,i}} = \frac{a w_{n-1,j}}{\sum_{i=1}^{n-1} a w_{n-1,i}} = \frac{w_{n-1,j}}{\sum_{i=1}^{n-1} w_{n-1,i}} \tag{135}
$$

We can use the expectation function to write down the naïve formula for the variance.

$$
\begin{aligned}
E_n(f(x)) &= E_{n-1}(f(x)) + \frac{w_{n,n}}{W_n}(f(x_n) - E_{n-1}(f(x))) && (136)\\
&= E_{n-1}(f(x)) + \alpha(f(x_n) - E_{n-1}(f(x))) && (137)\\
E_n(x^2) &= E_{n-1}(x^2) + \alpha(x_n^2 - E_{n-1}(x^2)) && (138)\\
\sigma_n^2 &= E_n(x^2) - \mu^2 && (139)
\end{aligned}
$$

So using the formula from the previous section we can write the incremental version:

$$
\begin{aligned}
S_n &= \frac{W_n - w_{n,n}}{W_{n-1}} S_{n-1} + w_{n,n}(x_n - \mu_n)(x_n - \mu_{n-1}) && (140)\\
S_n &= \frac{1-\alpha}{1} S_{n-1} + \alpha(x_n - \mu_n)(x_n - \mu_{n-1}) && (141)\\
\sigma_n^2 = \frac{S_n}{W_n} = S_n &= aS_{n-1} + (1-a)(x_n - \mu_n)(x_n - \mu_{n-1}) && (142)\\
&= (1-\alpha)(S_{n-1} + \alpha(x_n - \mu_{n-1})^2) && (143)
\end{aligned}
$$

This latter form is slightly more convenient for code:

```
diff := x - mean
incr := alpha * diff
mean := mean + incr
variance := (1 - alpha) * (variance + diff * incr)
```

# References


[1] Donald E. Knuth. *Seminumerical Algorithms*, volume 2 of *The Art of Computer Programming*, chapter 4.2.2, page 232. Addison-Wesley, Boston, third edition, 1998.

[2] Eric W. Weisstein. Sample variance computation. From Mathworld, a Wolfram web resource, `http://mathworld.wolfram.com/SampleVarianceComputation.html`.