Derivation of the Exponential Moving Average

R. Scott McIntire

Aug 30, 2024

1 Overview

The essence of the *exponential moving average* (EMA) is to provide a *localized* estimate of a time series as it proceeds. It does this by using a weighted average of the values in a "window" around each point of the series in way that each summand in the average is multiplied by powers of a factor, λ, in the interval $(0, 1)$. In what follows below we will assume a 1-based index, and that the data series $\mathbf{x}$ has length N.

Specifically, the exponential moving average (over a window of length m) is defined as:

$$e_n = \sum_{k=0}^{m-1} x_{n-k} w_{k+1} \tag{1}$$

To get the *exponential decay* in weights we assume that $w_{k+1} = \lambda w_k$.[1] Since the parameter λ is in the interval $(0, 1)$, it represents a decay factor. Notice that the indices of $\mathbf{x}$ and that of $\mathbf{w}$ go in the "opposite direction". This is done so that the largest weights are associated with the most recent data in the averaging window.

Before we go further, there is a problem with the formula as stated. If $n = 1$; or, in fact any value that is less than or equal to m, the expression x_{n-k} does not make sense for $k \geq n$. That is, at the start of the computation and for a little while, there is not a full window of values over a window of length m. As there is no information before x_1, we interpret these values to be x_1. We will apply this principal more generally below. That is, if $\mathbf{z}$ is a data series, we extend the definition of $\mathbf{z}$ so that it is defined for indices less than 1 as:

$$z_k = \begin{cases} z_k & k \geq 1; \\ z_1 & \text{otherwise.} \end{cases} \tag{2}$$

[1] The reason this is called exponential is that it is the discrete analog of continuous exponential decay. To see this let $f(x) = e^{-kt}$ for $k > 0$. Then sampling this function at regular intervals of say 1 we get the sequence: $\{e^{-k}, e^{-2k}, e^{-3k}, \ldots\}$. Here we can see that the elements of the sequence are powers of a value, $\lambda = e^{-k}$. That is, the sequence is nothing other than: $\{\lambda, \lambda^2, \lambda^3, \ldots\}$.

While we can use (1) to compute the EMA it is *inefficient*. The purpose of this paper is to derive a recursive formula that is fast to compute. The formula takes advantage of the fact that when the weights are exponentials (powers of λ) we don't have to recompute the bulk of the sums in the averaging window.

Regardless of the averaging scheme, we want our weights to sum to one.[2] But we also want the weights to be exponential; that is, $w_{k+1} = \lambda w_k$.

Ideally, all we want to do is give the decay factor, λ. If so, how do we determine the weights? The weights can be constructed as follows once a λ is given.

For $i \in [1, m]$ define

$$\hat{w}_i = \lambda^i \tag{3}$$

The $\hat{\mathbf{w}}$ have the property that they decay with λ; that is, $\lambda \hat{w}_k = \hat{w}_{k+1}$. However, how do we ensure that these values sum to 1 over the window?

To do this, we simply normalize the weights:

$$w_i = \frac{\hat{w}_i}{\sum_{k=1}^{m} \hat{w}_k} \quad \forall i \in [1, m] \tag{4}$$

Consequently,

$$\sum_{i=1}^{m} w_i = 1 \tag{5}$$

In doing this, did the proposed weights, $\mathbf{w}$, lose the decay factor property? The answer is no, the property remains:

$$\lambda w_i = w_{i+1} \quad \forall i \in [1, m-1]. \tag{6}$$

More generally, we can find *all* weights that have a decay factor of λ as they are solutions of the following difference equation:

$$w_{n+1} = \lambda\, w_n \tag{7}$$
$$w_1 = \alpha \tag{8}$$

The solution to this difference equation is:

$$w_n = \alpha\, \lambda^{n-1} \quad n \geq 1 \tag{9}$$

However, we need the weights to be *normalized* over a window of length, m:

$$\sum_{i=1}^{m} w_i = 1 \tag{10}$$

[2] If we consider the underlying data to be generated from a series of identically distributed random variables; then, treating the series elements as random variables we can view the moving average as an estimate of their mean, μ. To be a good estimate we would like this estimate to be *unbiased*: $\mathbb{E}\left[\sum_{k=1}^{m} x_{n-k} w_{k+1}\right] = \sum_{k=0}^{m-1} \mathbb{E}\left[x_{n-k}\right] w_{k+1}$. Which is $\mu \sum_{k=0}^{m-1} w_{k+1} = \mu \sum_{k=1}^{m} w_k = \mu$.

Plugging in our general formula for an exponential sequence we have:

$$\sum_{i=1}^{m} \alpha\lambda^{i-1} = 1 \tag{11}$$

$$\alpha\frac{1-\lambda^m}{1-\lambda} = 1 \tag{12}$$

Therefore, a decay factor, λ, and averaging window length, m, determine *the* initial weight, α, and thus determine *the* associated normalized weights:

$$\alpha = \frac{1-\lambda}{1-\lambda^m} \tag{13}$$

Note: We will use (2), (5), (6), and (13) in our derivations below.

2 Recursive Formula for EMA

We state again the definition of the exponential moving average of x as:

$$e_n = \sum_{k=0}^{m-1} x_{n-k} w_{k+1} \tag{14}$$

In order to find a recursive formula, we examine the next element in the smoothing sequence and try to relate it to the present element at index n.

$$\begin{aligned}
e_{n+1} &= \sum_{k=0}^{m-1} x_{n+1-k} w_{k+1} \\
&= \sum_{k=0}^{m-1} x_{n-(k-1)} w_{k+1} \\
&= \sum_{k=-1}^{m-2} x_{n-k} w_{k+2} \\
&= \sum_{k=0}^{m-2} x_{n-k} w_{k+2} + x_{n+1} w_1 \\
&= \lambda \sum_{k=0}^{m-2} x_{n-k} w_{k+1} + x_{n+1} w_1 \quad (\text{Since } \lambda w_{k+1} = w_{k+2}) \\
&= \lambda \sum_{k=0}^{m-1} x_{n-k} w_{k+1} + x_{n+1} w_1 - \lambda x_{n-(m-1)} w_m \\
&= \lambda \sum_{k=0}^{m-1} x_{n-k} w_{k+1} + x_{n+1} w_1 - \lambda x_{(n+1)-m} w_m
\end{aligned}$$

Finally, we may write[3]

$$
\begin{aligned}
e_{n+1} &= \lambda\left(e_n - x_{(n+1)-m}w_m\right) + \alpha x_{n+1} \quad \forall n \in [1, N-1] && (15)\\
e_1 &= x_1 && (16)
\end{aligned}
$$

The corresponding 0-based index formula is:[4]

$$
\begin{aligned}
e_n &= \lambda\left(e_{n-1} - x_{(n-m)}w_{m-1}\right) + \alpha x_n \quad \forall n \in [1, N-1] && (17)\\
e_0 &= x_0 && (18)
\end{aligned}
$$

3 Recursive Formula for the Std of EMA

In practice, one also wants to know a localized standard deviation of the process. In applications, one can provide "Bollinger" bands around the EMA estimate. These bands are (typically) some multiple of the standard deviation of the process.

We derive the standard deviation of the exponential moving average starting with the definition of the exponential moving average and its moving variation.[5]

$$
\begin{aligned}
e_n &= \sum_{k=0}^{m-1} x_{n-k}w_{k+1} && (19)\\
v_n &= \sum_{k=0}^{m-1} (x_{n-k} - e_{n-k})^2 w_{k+1} && (20)
\end{aligned}
$$

Note, that the form of the two equations is identical; that is, they both have the form:

$$
g_n = \sum_{k=0}^{m-1} f(x_{n-k} - c_{n-k})w_{k+1} \qquad (21)
$$

In the equation, (19), $c_{n-k} \equiv 0$; while in equation (20), $c_{n-k} = e_{n-k}$.

One can find a recursive formula g_n in *exactly* the same way as was done for (19). This is done by reworking the derivation of the recursive formula for (19), replacing x_{n-k} with $f(x_{n-k} - c_{n-k})$.

The recursive formula for g_n is

$$
\begin{aligned}
g_{n+1} &= \lambda\left(g_n - f(x_{(n+1)-m})w_m\right) + \alpha f(x_{n+1}) \quad \forall n \in [1, N-1] && (22)\\
g_1 &= f(x_1) && (23)
\end{aligned}
$$

[3] From the definition of e_n and using (2) (5), and (13) $e_1 = \sum_{k=0}^{m-1} x_{n-k}w_{k+1} = x_1 \sum_{k=0}^{m-1} w_{k+1} = x_1 \sum_{k=1}^{m} w_k = x_1$.

[4] We write the formulas down for both 1 and 0 based vector indices because computer languages typically use one or the other.

[5] This variance formula needs to be corrected by a factor which we leave to the end. The correction is a factor that is used to make the estimated standard deviation of the moving average *unbiased*. The formula for this factor is independent of the structure of the weights. This means that it can be pre-computed independently of the data series.

The 0-based index formula is:

$$
\begin{aligned}
g_n &= \lambda\left(g_{n-1} - f(x_{n-m})w_{m-1}\right) + \alpha f(x_n) \quad \forall n \in [1, N-1] && (24)\\
g_0 &= f(x_0) && (25)
\end{aligned}
$$

The corresponding f for (20) is $f(x) = x^2$; consequently, its recursive formula is

$$
\begin{aligned}
v_{n+1} &= \lambda\left(v_n - (x_{(n+1)-m} - e_{(n+1)-m})^2 w_m\right) + \alpha(x_{n+1} - e_{n+1})^2 \quad \forall n \in [1, N-1] && (26)\\
v_1 &= (x_1 - e_1)^2 = 0 && (27)
\end{aligned}
$$

The corresponding 0-based index formula is:

$$
\begin{aligned}
v_n &= \lambda\left(v_{n-1} - (x_{n-m} - e_{n-m})^2 w_{m-1}\right) + \alpha(x_n - e_n)^2 \quad \forall n \in [1, N-1] && (28)\\
v_0 &= (x_0 - e_0)^2 = 0 && (29)
\end{aligned}
$$

Finally, the standard deviation of the exponential moving average is:[6]

$$
s_n = \sqrt{\frac{v_n}{1 - \sum_{i=1}^{m} w_i^2}} \tag{30}
$$

The variance corrective factor[7], $1 - \sum_{i=1}^{m} w_i^2$, is a one time calculation independent of n. This factor is the standard correction when determining empirical variance of a weighted average.

Note: In practice, one may wish to use some prior estimate of the standard deviation of the process and smoothly move from that to the estimate above over the averaging windows when $n \in [1, m]$.

4 How to Specify the λ Factor

When using the EMA, practitioners do not specify the factor λ directly; instead they provide the "half-life", H, of the decay. By the half-file of λ, we mean the number of powers of λ, H, so that weight H is half the weight of weight 1. The condition for H is then:

$$
\frac{w_H}{w_1} = \frac{1}{2} \tag{31}
$$

$$
\frac{\alpha\lambda^{H-1}}{\alpha} = \lambda^{H-1} = \frac{1}{2} \tag{32}
$$

[6]The window length, m, must be greater than 1.

[7]This factor is derived in the appendix.

This can only be satisfied approximately for a given λ. [8]

In our case we are interested in the reverse; namely, given a half-life, H, determine the decay factor λ. To do this we solve the equation $\lambda^{H-1} = \frac{1}{2}$ for λ. Taking the log of this equation we have:[9]

$$\begin{aligned}
\log(\lambda^{H-1}) &= \log\left(\frac{1}{2}\right) \\
(H-1)\log(\lambda) &= \log(1) - \log(2) \\
\log(\lambda) &= -\frac{\log(2)}{H-1}
\end{aligned}$$

The formula to determine λ given the half-life H is:

$$\lambda = e^{-\frac{\log(2)}{H-1}} \tag{33}$$

For example, if the window of the moving average is 10, you might decide that the half-life occur around the middle of the window. Setting H to be 5, the decay factor, λ is $e^{-\frac{\log(2)}{4}} = 0.8408964152537145$.

[8]Note that H must be no more than the window length, m.
[9]The log function used below is the logarithm base e.

A Unbiased Estimator for a Sequence of IID Random Variables

In this section we derive the corrective formula for computing the empirical variance.

In what follows we assume that we are given a sequence of independent identically distributed random variables (IID): $\{x_i\}_{i=1}^N$ and corresponding weights: $\{w_i\}_{i=1}^N$. We assume that the random variables have a mean and a variance:

$$
\begin{aligned}
\mathbb{E}[x_i] &= \mu \quad \forall i, i \in [1, N] &(34)\\
\mathbb{E}[(x_i - \mu)]^2 &= \sigma^2 \quad \forall i, i \in [1, N] &(35)
\end{aligned}
$$

We assume that the weights are non-negative and are normalized:[10]

- $\sum_{i=1}^N w_i = 1$
- $\forall i, 0 \leq w_i < 1$

If we take the expectation of the weighted sum, $\sum_{i=1}^N x_i w_i$ we find that it is μ:[11]

$$
\mathbb{E}\left[\sum_{i=1}^{n} x_i w_i\right] = \sum_{i=1}^{N} \mathbb{E}[x_i n_i] = \sum_{i=1}^{N} \mathbb{E}[x_i] w_i = \mu \sum_{i=1}^{N} w_i = \mu \qquad (36)
$$

The weighted sum represents an estimate of the true mean of the IID random variables, which is μ. And in this case we see that it is a good estimate in that the expected value of the weighed combination of the random variables is μ.

We refer to such "good" estimates as *unbiased estimators*.

Now consider the sum $\sum_{i=1}^N (x_i - \mu)^2 w_i$. People use this as an estimate of the variance of the random variables. Is this a "good" (unbiased) estimate? Yes, it is as: $\mathbb{E}\left[\sum_{i=1}^N (x_i - \mu)^2 w_i\right] = \sum_{i=1}^N \mathbb{E}\left[(x_i - \mu)^2\right] w_i = \sigma^2 \sum_{i=1}^N w_i = \sigma^2$

But what if we replace the true mean, μ, above with our estimate: $\bar{x} = \sum_{i=1}^N x_i w_i$? In this case it turns out that the resulting estimate for the variance is *biased*. However, it happens that we are "off" by a certain factor and that factor depends only on the weights we use.[12]

To find this factor we examine the expected value of the straight forward formula one would imagine using to estimate the variance of the random variables.

[10] We also assume that the weight is not concentrated at any one index.

[11] This is due to the linearity of the expectation operator: $\mathbb{E}[ax + by + c] = a\mathbb{E}[x] + b\mathbb{E}[y] + c$, provided a, b, c are constants.

[12] In fact, the factor doesn't depend on the ordering of the weights.

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N}(x_i-\bar{x})^2 w_i\right] &= \sum_{i=1}^{N}\mathbb{E}\left[(x_i-\bar{x})^2\right] w_i \\
&= \sum_{i=1}^{N}\mathbb{E}\left[\left((x_i-\mu)+(\mu-\bar{x})\right)^2\right] w_i \\
&= \sum_{i=1}^{N}\mathbb{E}\left[(x_i-\mu)^2-2(x_i-\mu)(\bar{x}-\mu)+(\bar{x}-\mu)^2\right] w_i \\
&= \sum_{i=1}^{N}\left(\mathbb{E}\left[(x_i-\mu)^2\right]+\mathbb{E}\left[(\bar{x}-\mu)^2\right]-2\mathbb{E}\left[(x_i-\mu)(\bar{x}-\mu)\right]\right) w_i
\end{aligned} \tag{37}$$

We separately compute the expectation of each of the three summands.

$$\mathbb{E}\left[(x_i-\mu)^2\right] \quad = \quad \sigma^2 \tag{38}$$

$$
\begin{aligned}
\mathbb{E}\left[(\bar{x}-\mu)^2\right] &= \mathbb{E}\left[\left(\left(\sum_{j=1}^{N} x_j w_j\right)-\mu\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{j=1}^{N}(x_j w_j - \mu w_j)\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{j=1}^{N}(x_j - \mu) w_j\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{j=1}^{N}(x_j - \mu) w_j\right)\left(\sum_{k=1}^{N}(x_k - \mu) w_k\right)\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{N}\sum_{k=1}^{N}(x_j - \mu)(x_k - \mu) w_j w_k\right] \\
&= \left(\sum_{j=1}^{N}\sum_{k=1}^{N}\mathbb{E}\left[(x_j - \mu)(x_k - \mu)\right] w_j w_k\right) \\
&= \left(\sum_{j=1}^{N}\sum_{k=1}^{N}\sigma^2 w_j w_k \delta_{j,k}\right) \\
&= \left(\sum_{j=1}^{N}\sigma^2 w_j^2\right) \\
&= \sigma^2\left(\sum_{j=1}^{N} w_j^2\right)
\end{aligned} \tag{39}
$$

$$
\begin{aligned}
\mathbb{E}\left[(x_i-\mu)(\bar{x}-\mu)\right] &= \mathbb{E}\left[(x_i-\mu)\left(\left(\sum_{j=1}^{N} x_j w_j\right)-\mu\right)\right] \\
&= \mathbb{E}\left[(x_i-\mu)\left(\sum_{j=1}^{N}(x_j-\mu)w_j\right)\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{N}(x_i-\mu)(x_j-\mu)w_j\right] \\
&= \sum_{j=1}^{N}\mathbb{E}\left[(x_i-\mu)(x_j-\mu)\right]w_j \\
&= \sum_{j=1}^{N}\sigma^2 w_j \delta_{i,j} \\
&= \sigma^2 w_i
\end{aligned} \tag{40}
$$

Continuing with (39) we have

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N}(x_i-\bar{x})^2 w_i\right] &= \sum_{i=1}^{N}\mathbb{E}\left[(x_i-\bar{x})^2\right] w_i \\
&= \sum_{i=1}^{N}\mathbb{E}\left[\left((x_i-\mu)+(\mu-\bar{x})\right)^2\right] w_i \\
&= \sum_{i=1}^{N}\mathbb{E}\left[(x_i-\mu)^2 - 2(x_i-\mu)(\bar{x}-\mu) + (\bar{x}-\mu)^2\right] w_i \\
&= \sum_{i=1}^{N}\left(\mathbb{E}\left[(x_i-\mu)^2\right] + \mathbb{E}\left[\left(\sum_{j=1}^{N}(x_j-\mu)w_j\right)^2\right] - 2\mathbb{E}\left[(x_i-\mu)\left(\sum_{j=1}^{N}(x_j-\mu)w_j\right)\right]\right) w_i \\
&= \sum_{i=1}^{N}\left(\mathbb{E}\left[(x_i-\mu)^2\right] + \left(\sum_{j=1}^{N}\mathbb{E}\left[(x_j-\mu)^2\right] w_j^2\right) - 2\mathbb{E}\left[(x_i-\mu)\left(\sum_{j=1}^{N}(x_j-\mu)w_j\right)\right]\right) w_i \\
&= \sum_{i=1}^{N}\left(\sigma^2 + \sigma^2\sum_{j=1}^{N} w_j^2 - 2\left(\sum_{j=1}^{N}\mathbb{E}\left[(x_i-\mu)(x_j-\mu)\right] w_j\right)\right) w_i \\
&= \sum_{i=1}^{N}\left(\sigma^2 + \sigma^2\sum_{j=1}^{N} w_j^2 - 2\left(\sum_{j=1}^{N}\sigma^2\delta_{i,j} w_j\right)\right) w_i \\
&= \sigma^2\sum_{i=1}^{N}\left(1 + \sum_{j=1}^{N} w_j^2 - 2w_i\right) w_i \\
&= \sigma^2 + \sigma^2\left(\sum_{j=1}^{N} w_j^2\right) - 2\sigma^2\left(\sum_{i=1}^{N} w_i^2\right) \\
&= \sigma^2\left(1 - \sum_{i=1}^{N} w_i^2\right)
\end{aligned}
$$

Therefore, to make $\sum_{i=1}^{N}(x-\bar{x})^2 w_i$ an unbiased estimator of the variance σ^2 we need to correct by dividing by the factor $\left(1-\sum_{i=1}^{N} w_i^2\right)$.[13]

Consequently, $\frac{1}{1-\sum_{i=1}^{N} w_i^2}\sum_{i=1}^{N}(x_i-\bar{x})^2 w_i$ is an unbiased estimator for the variance of any of the IID variables: $\{x_i\}_{i=1}^{N}$. The unbiased standard deviation is then

$$
\sqrt{\frac{1}{1-\sum_{i=1}^{N} w_i^2}\sum_{i=1}^{N}(x_i-\bar{x})^2 w_i} \tag{41}
$$

[13] Note, this factor is non-zero when the weights are not concentrated at one point. As a consequence, the formula requires $N > 1$.

What is the formula when the weights are evenly distributed? That would mean that the weights are $w_i = \frac{1}{N}$. In this case the estimate for the standard deviation of the random variables is:

$$\begin{aligned}
\sqrt{\frac{1}{1-\sum_{i=1}^{N}\left(\frac{1}{N}\right)^2}\sum_{i=1}^{N}(x_i-\bar{x})^2\frac{1}{N}} &= \sqrt{\frac{1}{1-\frac{1}{N}}\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} \\
&= \sqrt{\frac{1}{\frac{N-1}{N}}\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} \\
&= \sqrt{\frac{N}{N-1}\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} \\
&= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2}
\end{aligned}$$

We see that we end up with the usual “sample” standard deviation.

A.1 Bias of Arbitrary Sample Moments.

We now proceed in the same way as above to compute the bias of more general moments.

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N}(x_i-\bar{x})^r w_i\right] &= \sum_{i=1}^{N}\mathbb{E}\left[(x_i-\bar{x})^r\right] w_i \qquad (42)\\
&= \sum_{i=1}^{N}\mathbb{E}\left[((x_i-\mu)+(\mu-\bar{x}))^r\right] w_i\\
&= \sum_{i=1}^{N}\mathbb{E}\left[\sum_{j=0}^{r}(-1)^j\binom{r}{j}(x_i-\mu)^{r-j}(\bar{x}-\mu)^j\right] w_i\\
&= \sum_{i=1}^{N}\sum_{j=0}^{r}(-1)^j\binom{r}{j}\mathbb{E}\left[(x_i-\mu)^{r-j}(\bar{x}-\mu)^j\right] w_i\\
&= \sum_{i=1}^{N}\sum_{j=0}^{r}(-1)^j\binom{r}{j}\mathbb{E}\left[(x_i-\mu)^{r-j}\left(\sum_{l=1}^{n}(x_l-\mu)w_l\right)^j\right] w_i\\
&= \sum_{i=1}^{N}\sum_{j=0}^{r}(-1)^j\binom{r}{j}\mathbb{E}\left[(x_i-\mu)^{r-j}\left((x_i-\mu)w_i+\sum_{\substack{l\in[1,N]\\ l\neq i}}(x_l-\mu)w_l\right)^j\right] w_i\\
&= \sum_{i=1}^{N}\sum_{j=0}^{r}(-1)^j\binom{r}{j}\mathbb{E}\left[(x_i-\mu)^{r-j}\sum_{k=0}^{j}\binom{j}{k}(x_i-\mu)^{j-k}w_i^{j-k}\left(\sum_{\substack{l\in[1,N]\\ l\neq i}}(x_l-\mu)w_l\right)^k\right] w_i\\
&= \sum_{i=1}^{N}\sum_{j=0}^{r}\sum_{k=0}^{j}(-1)^j\binom{r}{j}\binom{j}{k}\mathbb{E}\left[(x_i-\mu)^{r-k}w_i^{j-k}\left(\sum_{\substack{l\in[1,N]\\ l\neq i}}(x_l-\mu)w_l\right)^k\right] w_i\\
&= \sum_{i=1}^{N}\sum_{j=0}^{r}\sum_{k=0}^{j}(-1)^j\binom{r}{j}\binom{j}{k}\mathbb{E}\left[(x_i-\mu)^{r-k}\right]\mathbb{E}\left[\left(\sum_{\substack{l\in[1,N]\\ l\neq i}}(x_l-\mu)w_l\right)^k\right] w_i^{j+1-k}\\
&= \sum_{i=1}^{N}\sum_{j=0}^{r}(-1)^j\binom{r}{j}\mathbb{E}\left[(x_i-\mu)^r\right] w_i^{j+1}\\
&\quad + \sum_{i=1}^{N}\sum_{j=2}^{r}\sum_{k=2}^{j}(-1)^j\binom{r}{j}\binom{j}{k}\mathbb{E}\left[(x_i-\mu)^{r-k}\right]\mathbb{E}\left[\left(\sum_{\substack{l\in[1,N]\\ l\neq i}}(x_l-\mu)w_l\right)^k\right] w_i^{j+1-k}
\end{aligned}
$$

Consequently,

$$\mathbb{E}\left[\sum_{i=1}^{N}(x_i-\bar{x})^r w_i\right] = M_r + M_r\sum_{j=1}^{r}(-1)^j\binom{r}{j}\sum_{i=1}^{N}w_i^{j+1} + \sum_{i=1}^{N}\sum_{j=2}^{r}\sum_{k=2}^{j}(-1)^j\binom{r}{j}\binom{j}{k}M_{r-k} \tag{43}$$

$$\times\,\mathbb{E}\left[\sum_{\substack{k_1+k_2+\cdots+k_N=k\\ \forall n, k_n\geq 0\\ k_i=0}}\binom{k}{k_1,k_2,\ldots,k_N}\prod_{\substack{l\in[1,N]\\ l\neq i}}(x_l-\mu)^{k_l}w_l^{k_l}\right]w_i^{j+1-k}$$

$$= M_r + M_r\sum_{j=1}^{r}(-1)^j\binom{r}{j}\sum_{i=1}^{N}w_i^{j+1} + \sum_{i=1}^{N}\sum_{j=2}^{r}\sum_{k=2}^{j}(-1)^j\binom{r}{j}\binom{j}{k}M_{r-k}$$

$$\times\sum_{\substack{k_1+k_2+\cdots+k_N=k\\ \forall n, k_n\geq 0\,\wedge\, k_n\neq 1\\ k_i=0}}\binom{k}{k_1,k_2,\ldots,k_N}\left(\prod_{\substack{l\in[1,N]\\ l\neq i}}\mathbb{E}\left[(x_l-\mu)^{k_l}\right]w_l^{k_l}\right)w_i^{j+1-k}$$

$$= M_r + M_r\sum_{j=1}^{r}(-1)^j\binom{r}{j}\sum_{i=1}^{N}w_i^{j+1} + \sum_{i=1}^{N}\sum_{j=2}^{r}\sum_{k=2}^{j}(-1)^j\binom{r}{j}\binom{j}{k}M_{r-k}$$

$$\times\sum_{\substack{k_1+k_2+\cdots+k_N=k\\ \forall n, k_n\geq 0\,\wedge\, k_n\neq 1\\ k_i=0}}\binom{k}{k_1,k_2,\ldots,k_N}\left(\prod_{l=1}^{N}M_{k_l}w_l^{k_l}\right)w_i^{j+1-k}$$

Here, we use the notation, M_r, for the r^{th} moment of the distribution of the $\{x_i\}_{i=1}^N$.

We break up the multinomial formula into a sum with respect to k_i producing the final formula for the expected value of the naïve r^{th} moment:[14]

$$\mathbb{E}\left[(x_i-\bar{x})^r w_i\right] = M_r + M_r\sum_{j=1}^{r}(-1)^j\binom{r}{j}\sum_{i=1}^{N}w_i^{j+1} + \sum_{j=2}^{r}\sum_{k=2}^{j}(-1)^j\binom{r}{j}\binom{j}{k}M_{r-k} \tag{44}$$

$$\times\left[\sum_{i=1}^{N}w_i^{j+1-k}\left(\sum_{\substack{k_1+k_2+\cdots+k_N=k\\ \forall n, k_n\geq 0\,\wedge\, k_n\neq 1}}\binom{k}{k_1,k_2,\ldots,k_N}\prod_{l=1}^{N}M_{k_l}w_l^{k_l}\right)\right.$$

$$\left.-\sum_{z=2}^{k}M_z\left(\sum_{i=1}^{N}w_i^{i+z+1-k}\sum_{\substack{k_1+k_2+\cdots+k_{N-1}=k-z\\ \forall n, k_n\geq 0\,\wedge\, k_n\neq 1}}\binom{k-z}{k_1,k_2,\ldots,k_{N-1}}\prod_{\substack{l\in[1,N-1]\\ l\neq i}}M_{k_l}w_l^{k_l}\right)\right]$$

[14] We use the standard interpretation of any sum of the form, $\sum_{i=n}^{m} f(i)$, to be 0 if $m < n$.

We check the case $r = 2$ to see if it coincides with the previous section:

$$\mathbb{E}\left[(x_i - \bar{x})^2 w_i\right] = \sigma^2 + \sigma^2 \sum_{j=1}^{2} (-1)^j \binom{2}{j} \sum_{i=1}^{N} w_i^{j+1} + \sum_{j=2}^{2} \sum_{k=2}^{j} (-1)^j \binom{2}{j} \binom{j}{k} M_{2-k} \tag{45}$$
$$\times \sum_{i=1}^{N} \left[\sum_{\substack{k_1+k_2+\cdots+k_N=k \\ \forall n, k_n \geq 0 \wedge k_n \neq 1}} \binom{k}{k_1, k_2, \ldots, k_N} \left(\prod_{l=1}^{N} M_{k_l} w_l^{k_l} \right) w_i^{j+1-k} \right.$$
$$\left. - \sum_{z=2}^{k} M_z \sum_{\substack{k_1+k_2+\cdots+k_{N-1}=k-z \\ \forall n, k_n \geq 0 \wedge k_n \neq 1}} \binom{k-z}{k_1, k_2, \ldots, k_{N-1}} \left(\prod_{\substack{l \in [1, N-1] \\ l \neq i}} M_{k_l} w_l^{k_l} \right) w_i^{j+z+1-k} \right]$$

This becomes

$$\mathbb{E}\left[(x_i - \bar{x})^2 w_i\right] = \sigma^2 + \sigma^2 \sum_{j=1}^{2} (-1)^j \binom{2}{j} \sum_{i=1}^{N} w_i^{j+1} \tag{46}$$
$$+ \sum_{\substack{k_1+k_2+\cdots+k_N=2 \\ \forall n, k_n \geq 0 \wedge k_n \neq 1}} \binom{2}{k_1, k_2, \ldots, k_N} \left(\prod_{l=1}^{N} M_{k_l} w_l^{k_l} \right) \left(\sum_{i=1}^{N} w_i \right) - \sigma^2 \left(\sum_{i=1}^{N} w_i^3 \right)$$

Which is

$$\mathbb{E}\left[(x_i - \bar{x})^2 w_i\right] = \sigma^2 + \sigma^2 \left(-2 \sum_{i=1}^{N} w_i^2 + \sum_{i=1}^{N} w_i^3 \right) + \sigma^2 \sum_{l=1}^{N} w_l^2 \left(\sum_{i=1}^{N} w_i \right) - \sigma^2 \left(\sum_{i=1}^{N} w_i^3 \right) \tag{47}$$

This becomes

$$\mathbb{E}\left[(x_i - \bar{x})^2 w_i\right] = \sigma^2 \left(1 - \sum_{i=1}^{N} w_i^2 \right) \tag{48}$$

As a more complicated example we try $r = 3$. In this case the formula is:

$$\mathbb{E}\left[(x_i - \bar{x})^3 w_i\right] = M_3 + M_3 \sum_{j=1}^{3} (-1)^j \binom{3}{j} \sum_{i=1}^{N} w_i^{j+1} + \sum_{j=2}^{3} \sum_{k=2}^{j} (-1)^j \binom{3}{j} \binom{j}{k} M_{3-k} \tag{49}$$
$$\times \sum_{i=1}^{N} \left[\sum_{\substack{k_1+k_2+\cdots+k_N=k \\ \forall n, k_n \geq 0 \wedge k_n \neq 1}} \binom{k}{k_1, k_2, \ldots, k_N} \left(\prod_{l=1}^{N} M_{k_l} w_l^{k_l} \right) w_i^{j+1-k} \right.$$
$$\left. - \sum_{z=2}^{k} M_z \sum_{\substack{k_1+k_2+\cdots+k_{N-1}=k-z \\ \forall n, k_n \geq 0 \wedge k_n \neq 1}} \binom{k-z}{k_1, k_2, \ldots, k_{N-1}} \left(\prod_{\substack{l \in [1, N-1] \\ l \neq i}} M_{k_l} w_l^{k_l} \right) w_i^{j+z+1-k} \right]$$

Which is

$$\mathbb{E}\left[(x_i - \bar{x})^3 w_i\right] = M_3 + M_3 \left(-3 \sum_{i=1}^{2} w_i^2 + 3 \sum_{i=1}^{N} w_i^3 - \sum_{i=1}^{N} w_i^4 \right) - M_3 \left(\sum_{l=1}^{N} w_l^3 \left(\sum_{i=1}^{N} w_i \right) - \sum_{i=1}^{N} w_i^4 \right) \tag{50}$$

Or,

$$\mathbb{E}\left[(x_i - \bar{x})^3 w_i\right] = M_3 \left(1 - 3 \sum_{i=1}^{N} w_i^2 + 2 \sum_{i=1}^{N} w_i^3 \right) \tag{51}$$

Therefore, the corresponding unbiased estimator for the 3^{rd} moment is:[15]

$$\frac{\sum_{i=1}^{N} (x_i - \bar{x})^3 w_i}{1 - 3 \sum_{i=1}^{N} w_i^2 + 2 \sum_{i=1}^{N} w_i^3} \tag{52}$$

In the more general case, $r > 3$, there may be more work to determine a corrective procedure to render the naïve r^{th} moment unbiased.

As a last example we try the case $r = 4$. The naïve estimate for the 4^{th} moment is:

[15] As mentioned in the case of the second moment, this formula does not apply when the weights are concentrated at one point. Consequently, we must have $N > 1$.

$$
\begin{aligned}
\mathbb{E}\left[(x_i-\bar{x})^4 w_i\right] = {} & M_4 + M_4 \sum_{j=1}^{4} (-1)^j \binom{4}{j} \sum_{i=1}^{N} w_i^{j+1} + \sum_{j=2}^{4} \sum_{k=2}^{j} (-1)^j \binom{4}{j} \binom{j}{k} M_{4-k} \\
& \times \sum_{i=1}^{N} \left[\sum_{\substack{k_1+k_2+\cdots+k_N=k \\ \forall n, k_n \geq 0 \wedge k_n \neq 1}} \binom{k}{k_1, k_2, \ldots, k_N} \left(\prod_{l=1}^{N} M_{k_l} w_l^{k_l} \right) w_i^{j+1-k} \right. \\
& \left. - \sum_{z=2}^{k} M_z \sum_{\substack{k_1+k_2+\cdots+k_{N-1}=k-z \\ \forall n, k_n \geq 0 \wedge k_n \neq 1}} \binom{k-z}{k_1, k_2, \ldots, k_{N-1}} \left(\prod_{\substack{l \in [1, N-1] \\ l \neq i}} M_{k_l} w_l^{k_l} \right) w_i^{j+z+1-k} \right]
\end{aligned} \tag{53}
$$

This is

$$
\begin{aligned}
\mathbb{E}\left[(x_i-\bar{x})^4 w_i\right] = {} & M_4 \left(1 - 4 \sum_{i=1}^{N} w_i^2 + 6 \sum_{i=1}^{N} w_i^3 - 4 \sum_{i=1}^{N} w_i^4 + \sum_{i=1}^{N} w_i^5 \right) \\
& + \binom{4}{2}\binom{2}{2} M_2^2 \left(\sum_{l=1}^{N} w_l^2 \left(\sum_{i=1}^{N} w_i \right) - \sum_{j=1}^{N} w_l^2 \left(\sum_{i=1}^{N} w_i^3 \right) \right) \\
& - \binom{4}{3}\binom{3}{2} M_2^2 \left(\sum_{l=1}^{N} w_l^2 \left(\sum_{i=1}^{N} w_i^2 \right) - \sum_{j=1}^{N} w_l^2 \left(\sum_{i=1}^{N} w_i^4 \right) \right) \\
& + \binom{4}{4}\binom{4}{2} M_2^2 \left(\sum_{l=1}^{N} w_l^2 \left(\sum_{i=1}^{N} w_i^3 \right) - \sum_{j=1}^{N} w_l^2 \left(\sum_{i=1}^{N} w_i^5 \right) \right) \\
& + \binom{4}{4}\binom{4}{4} M_4 \sum_{i=1}^{N} \left(\sum_{l=1}^{N} w_l^4 w_i \right) \\
& + \binom{4}{2,2} M_2^2 \sum_{l=1}^{N-1} \sum_{m=l+1}^{N} w_l^2 w_m^2 \left(\sum_{i=1}^{N} w_i \right) \\
& - M_2 \sum_{i=1}^{N} \left(\sum_{\substack{l \in [1, N] \\ l \neq i}} M_2 w_l^2 \right) w_i^3 - M_4 \sum_{i=1}^{N} w_i^5
\end{aligned} \tag{54}
$$

$$
\begin{aligned}
\mathbb{E}\left[(x_i-\bar{x})^4 w_i\right] = {} & M_4\left(1-3\sum_{i=1}^{N} w_i^2+6\sum_{i=1}^{N} w_i^3-3\sum_{i=1}^{N} w_i^4+\sum_{i=1}^{N} w_i^5-\sum_{i=1}^{N} w_i^5\right) \\
& +6M_2^2\left(\sum_{i=1}^{N} w_i^2-\left(\sum_{i=1}^{N} w_i^2\right)\left(\sum_{i=1}^{N} w_i^3\right)\right) \\
& -12M_2^2\left(\left(\sum_{i=1}^{N} w_i^2\right)\left(\sum_{i=1}^{N} w_i^2\right)-\left(\sum_{i=1}^{N} w_i^2\right)\left(\sum_{i=1}^{N} w_i^4\right)\right) \\
& +6M_2^2\left(\left(\sum_{i=1}^{N} w_i^2\right)\left(\sum_{i=1}^{N} w_i^3\right)-\left(\sum_{i=1}^{N} w_i^2\right)\left(\sum_{i=1}^{N} w_i^5\right)\right) \\
& +6M_2^2\sum_{l=1}^{N-1}\sum_{m=l+1}^{N} w_l^2 w_m^2 \\
& -M_2^2\sum_{i=1}^{N}\left(\sum_{\substack{l\in[1,N]\\ l\neq i}} w_l^2\right) w_i^3
\end{aligned}
\tag{55}
$$

Let $W_j=\sum_{i=1}^{N} w_i^j$ for $j\in[1,5]$. We may then write the above as:

$$
\begin{aligned}
\mathbb{E}\left[(x_i-\bar{x})^4 w_i\right] = {} & M_4\left(1-3W_2+6W_3-3W_4\right) \\
& +6M_2^2W_2-12M_2^2W_2^2+12M_2^2W_2W_4-6M_2^2W_2W_5 \\
& +6M_2^2\sum_{l=1}^{N-1}\sum_{m=l+1}^{N} w_l^2 w_m^2-M_2^2\sum_{i=1}^{N}\sum_{\substack{l\in[1,N]\\ l\neq i}} w_l^2 w_i^3
\end{aligned}
\tag{56}
$$

Or,

$$
\begin{aligned}
\mathbb{E}\left[(x_i-\bar{x})^4 w_i\right] = {} & M_4\left(1-3W_2+6W_3-3W_4\right) \\
& +6M_2^2W_2-12M_2^2W_2^2+12M_2^2W_2W_4-6M_2^2W_2W_5 \\
& +6M_2^2\sum_{l=1}^{N-1}\sum_{m=l+1}^{N} w_l^2 w_m^2-M_2^2W_2W_3+M_2^2W_5
\end{aligned}
\tag{57}
$$

The relative 4^{th} naïve moment becomes:

$$\frac{\mathbb{E}\left[(x_i-\bar{x})^4 w_i\right]}{M_2^2} = \frac{M_4}{M_2^2}\left(1-3W_2+6W_3-3W_4\right) \\ +6W_2-12W_2^2+12W_2W_4-6W_2W_5 \\ +6\sum_{l=1}^{N-1}\sum_{m=l+1}^{N} w_l^2 w_m^2 - W_2W_3 + W_5 \tag{58}$$

Therefore, the estimate for the unbiased relative 4^{th} moment is:

$$\frac{M_4}{M_2^2} = \frac{\frac{\mathbb{E}\left[(x_i-\bar{x})^4 w_i\right]}{M_2^2} - 6W_2 + 12W_2^2 - 12W_2W_4 + 6W_2W_5 - 6\sum_{l=1}^{N-1}\sum_{m=l+1}^{N} w_l^2 w_m^2 + W_2W_3 - W_5}{(1-3W_2+6W_3-3W_4)} \tag{59}$$

Since we don't know $M_2=\sigma^2$, we use our unbiased estimate: $\frac{\sum_{j=1}^{N}(x-\bar{x})^2 w_i}{1-\sum_{i=1}^{N} w_i^2}$. We also replace $\mathbb{E}\left[\sum_{i=1}^{N}(x_i-\bar{x})^4 w_i\right]$ with the sample estimate: $\sum_{i=1}^{N}(x_i-\bar{x})^4 w_i$.

Therefore, our unbiased estimate of the relative Kurtosis is:

$$\frac{M_4}{M_2^2} = \frac{\frac{(1-W_2)^2\sum_{i=1}^{N}(x_i-\bar{x})^4 w_i}{\left(\sum_{i=1}^{N}(x_i-\bar{x})^2 w_i\right)^2} - 6W_2 + 12W_2^2 - 12W_2W_4 + 6W_2W_5 - 6\sum_{l=1}^{N-1}\sum_{m=l+1}^{N} w_l^2 w_m^2 + W_2W_3 - W_5}{(1-3W_2+6W_3-3W_4)} \tag{60}$$