Computing Moments with fromo

Steven E. Pav *

November 29, 2024

Abstract

The fromo package provides fast robust summation using the Welford-Terriberry method. The update formula used therein is described, as well as the output produced by various functions.

1 The update formula

Let $\mathcal{A}$ be a set of indices over the data x_i and corresponding weights $w_i > 0$. The weights are *replication weights*, and are intended to simulate having observed multiple identical, though independent, values of x_i. In the standard setting the weights are identically 1.

Define the total elements, sum of weights, and (weighted) mean over $\mathcal{A}$ via

$$\begin{aligned} n_{\mathcal{A}} &= \left|\mathcal{A}\right|, \\ W_{\mathcal{A}} &= \sum_{i\in\mathcal{A}} w_i, \\ \mu_{\mathcal{A}} &= \frac{\sum_{i\in\mathcal{A}} w_i x_i}{W_{\mathcal{A}}}. \end{aligned}$$

Then go on to define the k^{th} centered weighted sum via

$$\mathcal{S}_{\mathcal{A},k} = \sum_{i\in\mathcal{A}} w_i \left(x_i - \mu_{\mathcal{A}}\right)^k.$$

Note that we have

$$\mathcal{S}_{\mathcal{A},0} = W_{\mathcal{A}}, \qquad \text{and} \qquad \mathcal{S}_{\mathcal{A},1} = 0.$$

When $\mathcal{A}$ consists of a single observation, that is when $n_{\mathcal{A}} = 1$, we have $\mu_{\mathcal{A}} = x_a$ for the unique $a \in \mathcal{A}$, and $\mathcal{S}_{\mathcal{A},k} = 0$ for all $k \geq 1$.

Let $\mathcal{B}$ and $\mathcal{C}$ be sets of indices with the restriction that $\mathcal{A} \cap \mathcal{B} = \emptyset$, $\mathcal{C} \subseteq \mathcal{A}$, and define

$$\mathcal{D} = \mathcal{A} \cup \mathcal{B} \setminus \mathcal{C}.$$

Thus $\mathcal{D}$ is $\mathcal{A}$ 'plus' $\mathcal{B}$ 'minus' $\mathcal{C}$. Consider how to compute $W_{\mathcal{D}}$, $\mu_{\mathcal{D}}$, and $\mathcal{S}_{\mathcal{D},k}$ from the sums of weights, weighted means, and centered sums over the sets $\mathcal{A}$,

*shabbychef@gmail.com

$\mathcal{B}$, and $\mathcal{C}$. [1, 2, 4] We have:

$$\begin{aligned}
n_{\mathcal{D}} &= n_{\mathcal{A}} + n_{\mathcal{B}} - n_{\mathcal{C}}, \\
W_{\mathcal{D}} &= W_{\mathcal{A}} + W_{\mathcal{B}} - W_{\mathcal{C}}, \\
\mu_{\mathcal{D}} &= \frac{W_{\mathcal{A}}\mu_{\mathcal{A}} + W_{\mathcal{B}}\mu_{\mathcal{B}} - W_{\mathcal{C}}\mu_{\mathcal{C}}}{W_{\mathcal{D}}}, \\
&= \mu_{\mathcal{A}} + \frac{W_{\mathcal{B}}\left(\mu_{\mathcal{B}} - \mu_{\mathcal{A}}\right) - W_{\mathcal{C}}\left(\mu_{\mathcal{C}} - \mu_{\mathcal{A}}\right)}{W_{\mathcal{D}}},
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{S}_{\mathcal{D},k} &= \sum_{i\in\mathcal{D}} w_i \left(x_i - \mu_{\mathcal{D}}\right)^k, \\
&= \sum_{i\in\mathcal{A}} w_i \left(x_i - \mu_{\mathcal{D}}\right)^k + \sum_{i\in\mathcal{B}} w_i \left(x_i - \mu_{\mathcal{D}}\right)^k - \sum_{i\in\mathcal{C}} w_i \left(x_i - \mu_{\mathcal{D}}\right)^k, \\
&= \sum_{i\in\mathcal{A}} w_i \left(x_i - \mu_{\mathcal{A}} + \mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^k + \sum_{i\in\mathcal{B}} w_i \left(x_i - \mu_{\mathcal{B}} + \mu_{\mathcal{B}} - \mu_{\mathcal{D}}\right)^k - \sum_{i\in\mathcal{C}} w_i \left(x_i - \mu_{\mathcal{C}} + \mu_{\mathcal{C}} - \mu_{\mathcal{D}}\right)^k, \\
&= \sum_{i\in\mathcal{A}} \sum_{0\le j\le k} \binom{k}{j} w_i \left(x_i - \mu_{\mathcal{A}}\right)^j \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^{k-j} \\
&\quad + \sum_{i\in\mathcal{B}} \sum_{0\le j\le k} \binom{k}{j} w_i \left(x_i - \mu_{\mathcal{B}}\right)^j \left(\mu_{\mathcal{B}} - \mu_{\mathcal{D}}\right)^{k-j} \\
&\quad - \sum_{i\in\mathcal{C}} \sum_{0\le j\le k} \binom{k}{j} w_i \left(x_i - \mu_{\mathcal{C}}\right)^j \left(\mu_{\mathcal{C}} - \mu_{\mathcal{D}}\right)^{k-j}, \\
&= \sum_{0\le j\le k} \binom{k}{j} \left\{ \mathcal{S}_{\mathcal{A},j} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^{k-j} + \mathcal{S}_{\mathcal{B},j} \left(\mu_{\mathcal{B}} - \mu_{\mathcal{D}}\right)^{k-j} - \mathcal{S}_{\mathcal{C},j} \left(\mu_{\mathcal{C}} - \mu_{\mathcal{D}}\right)^{k-j} \right\}, \\
&= \mathcal{S}_{\mathcal{A},k} + \mathcal{S}_{\mathcal{B},k} - \mathcal{S}_{\mathcal{C},k} + W_{\mathcal{A}} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^k + W_{\mathcal{B}} \left(\mu_{\mathcal{B}} - \mu_{\mathcal{D}}\right)^k - W_{\mathcal{C}} \left(\mu_{\mathcal{C}} - \mu_{\mathcal{D}}\right)^k \\
&\quad + \sum_{2\le j<k} \binom{k}{j} \left\{ \mathcal{S}_{\mathcal{A},j} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^{k-j} + \mathcal{S}_{\mathcal{B},j} \left(\mu_{\mathcal{B}} - \mu_{\mathcal{D}}\right)^{k-j} - \mathcal{S}_{\mathcal{C},j} \left(\mu_{\mathcal{C}} - \mu_{\mathcal{D}}\right)^{k-j} \right\}.
\end{aligned}$$

Note that if the centered sums are to be computed *in place*, that is, overwriting the vector of $\mathcal{S}_{\mathcal{A},j}$ with values of $\mathcal{S}_{\mathcal{D},j}$, then they should be computed in decreasing order, as updates to higher order sums require the old values of lower order sums.

It should also be noted that we did not use the restriction that $w_i > 0$. True, negative values of w_i can cause $W_{\mathcal{D}}$ to be zero, and therefore $\mu_{\mathcal{D}}$ is not defined, nor is $\mathcal{S}_{\mathcal{D},k}$. Negative weights also make no sense for most statistical uses. However, notice what happens when we subsitute every w_c with $-w_c$ for $c \in \mathcal{C}$. If we compute the total weight, $W_{\mathcal{C}}$, its sign has flipped, as has that of $\mathcal{S}_{\mathcal{C},k}$, but not of $\mu_{\mathcal{C}}$. In this case, using the negative weights we have

$$\begin{aligned}
\mathcal{S}_{\mathcal{D},k} &= \mathcal{S}_{\mathcal{A},k} + \mathcal{S}_{\mathcal{B},k} + \tilde{\mathcal{S}}_{\mathcal{C},k} + W_{\mathcal{A}} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^k + W_{\mathcal{B}} \left(\mu_{\mathcal{B}} - \mu_{\mathcal{D}}\right)^k + \tilde{W}_{\mathcal{C}} \left(\tilde{\mu}_{\mathcal{C}} - \mu_{\mathcal{D}}\right)^k \\
&\quad + \sum_{2\le j<k} \binom{k}{j} \left\{ \mathcal{S}_{\mathcal{A},j} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^{k-j} + \mathcal{S}_{\mathcal{B},j} \left(\mu_{\mathcal{B}} - \mu_{\mathcal{D}}\right)^{k-j} + \tilde{\mathcal{S}}_{\mathcal{C},j} \left(\tilde{\mu}_{\mathcal{C}} - \mu_{\mathcal{D}}\right)^{k-j} \right\},
\end{aligned}$$

where the $\tilde{W}_{\mathcal{C}}$, $\tilde{\mu}_{\mathcal{C}}$, and $\tilde{\mathcal{S}}_{\mathcal{C},j}$ denote that they are computed with negative weights. Now the update formula for removing $\mathcal{C}$ from $\mathcal{A}$ looks just like that for adding $\mathcal{B}$, except negative weights are used.

1.1 Adding a single observation

We now consider the special case where $\mathcal{C}$ is empty, and $\mathcal{B}$ consists of the single index b. We then have $W_{\mathcal{D}} = W_{\mathcal{A}} + w_b$,

$$
\begin{aligned}
\mu_{\mathcal{D}} &= \frac{W_{\mathcal{A}}\mu_{\mathcal{A}} + w_b x_b}{W_{\mathcal{A}} + w_b}, \\
&= \mu_{\mathcal{A}} + w_b \frac{x_b - \mu_{\mathcal{A}}}{W_{\mathcal{A}} + w_b}, \\
&= \mu_{\mathcal{A}} + w_b \frac{x_b - \mu_{\mathcal{A}}}{W_{\mathcal{D}}}.
\end{aligned}
$$

So then

$$
\mu_{\mathcal{A}} - \mu_{\mathcal{D}} = -w_b \frac{x_b - \mu_{\mathcal{A}}}{W_{\mathcal{D}}}.
$$

Also

$$
\begin{aligned}
x_b - \mu_{\mathcal{D}} &= x_b - \mu_{\mathcal{A}} + \mu_{\mathcal{A}} - \mu_{\mathcal{D}}, \\
&= x_b - \mu_{\mathcal{A}} - w_b \frac{x_b - \mu_{\mathcal{A}}}{W_{\mathcal{D}}}, \\
&= W_{\mathcal{A}} \frac{x_b - \mu_{\mathcal{A}}}{W_{\mathcal{D}}}.
\end{aligned}
$$

Since $\mathcal{B}$ is a single index, $\mathcal{S}_{\mathcal{B},k} = 0$ for $k > 0$, then we have

$$
\begin{aligned}
\mathcal{S}_{\mathcal{D},k} &= \mathcal{S}_{\mathcal{A},k} + W_{\mathcal{A}} \left(-w_b \frac{x_b - \mu_{\mathcal{A}}}{W_{\mathcal{D}}} \right)^k + w_b \left(W_{\mathcal{A}} \frac{x_b - \mu_{\mathcal{A}}}{W_{\mathcal{D}}} \right)^k \\
&\quad + \sum_{2 \le j < k} \binom{k}{j} \mathcal{S}_{\mathcal{A},j} \left(-w_b \frac{x_b - \mu_{\mathcal{A}}}{W_{\mathcal{D}}} \right)^{k-j}, \\
&= \mathcal{S}_{\mathcal{A},k} + w_b \left(x_b - \mu_{\mathcal{D}} \right)^k \left[1 + \left(\frac{-w_b}{W_{\mathcal{A}}} \right)^{k-1} \right] \\
&\quad + \sum_{2 \le j < k} \binom{k}{j} \mathcal{S}_{\mathcal{A},j} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}} \right)^{k-j}, \\
&= \mathcal{S}_{\mathcal{A},k} + W_{\mathcal{A}} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}} \right)^k \left[1 - \left(\frac{W_{\mathcal{A}}}{-w_b} \right)^{k-1} \right] \\
&\quad + \sum_{2 \le j < k} \binom{k}{j} \mathcal{S}_{\mathcal{A},j} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}} \right)^{k-j}.
\end{aligned}
$$

For the $k = 2$ case this further simplifies to

$$
\begin{aligned}
\mathcal{S}_{\mathcal{D},2} &= \mathcal{S}_{\mathcal{A},2} + \frac{W_{\mathcal{A}} w_b}{W_{\mathcal{D}}} \left(x_b - \mu_{\mathcal{A}} \right)^2, \\
&= \mathcal{S}_{\mathcal{A},2} + \frac{W_{\mathcal{D}} w_b}{W_{\mathcal{A}}} \left(x_b - \mu_{\mathcal{D}} \right)^2, \\
&= \mathcal{S}_{\mathcal{A},2} + W_{\mathcal{D}} \left(x_b - \mu_{\mathcal{D}} \right) \left(\mu_{\mathcal{D}} - \mu_{\mathcal{A}} \right).
\end{aligned}
$$

Note that the quantity $W_{\mathcal{D}} \left(\mu_{\mathcal{D}} - \mu_{\mathcal{A}} \right) = w_b \left(x_b - \mu_{\mathcal{A}} \right)$ is computed as an intermediary in updating the weighted mean, resulting in some computational savings.

1.2 Removing a single observation

As noted above, removing a single observation should look just like adding a single observation, except signs on the weights and weighted sums are flipped. Let c be the single index in $\mathcal{C}$. We then have $W_{\mathcal{D}} = W_{\mathcal{A}} - w_c$, and

$$\mu_{\mathcal{D}} = \mu_{\mathcal{A}} - w_c \frac{x_c - \mu_{\mathcal{A}}}{W_{\mathcal{D}}}.$$

The centered sum update formula is

$$\begin{aligned}
\mathcal{S}_{\mathcal{D},k} &= \mathcal{S}_{\mathcal{A},k} - w_c \left(x_c - \mu_{\mathcal{D}}\right)^k \left[1 + \left(\frac{w_c}{W_{\mathcal{A}}}\right)^{k-1}\right] \\
&\quad + \sum_{2 \le j < k} \binom{k}{j} \mathcal{S}_{\mathcal{A},j} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^{k-j}, \\
&= \mathcal{S}_{\mathcal{A},k} + W_{\mathcal{A}} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^k \left[1 - \left(\frac{W_{\mathcal{A}}}{w_c}\right)^{k-1}\right] \\
&\quad + \sum_{2 \le j < k} \binom{k}{j} \mathcal{S}_{\mathcal{A},j} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^{k-j}.
\end{aligned}$$

For the $k = 2$ case this further simplifies to

$$\begin{aligned}
\mathcal{S}_{\mathcal{D},2} &= \mathcal{S}_{\mathcal{A},2} - \frac{W_{\mathcal{A}} w_c}{W_{\mathcal{D}}} \left(x_c - \mu_{\mathcal{A}}\right)^2, \\
&= \mathcal{S}_{\mathcal{A},2} - \frac{W_{\mathcal{D}} w_c}{W_{\mathcal{A}}} \left(x_c - \mu_{\mathcal{D}}\right)^2, \\
&= \mathcal{S}_{\mathcal{A},2} + W_{\mathcal{D}} \left(x_c - \mu_{\mathcal{D}}\right) \left(\mu_{\mathcal{D}} - \mu_{\mathcal{A}}\right).
\end{aligned}$$

1.3 Adding and removing a single observation

Consider the case of adding a single $b \in \mathcal{B}$ and removing a single $c \in \mathcal{C}$. The total number elements, $n_{\mathcal{A}}$, is unchanged, of course. We have

$$\begin{aligned}
W_{\mathcal{D}} &= W_{\mathcal{A}} + w_b - w_c, \\
\mu_{\mathcal{D}} &= \mu_{\mathcal{A}} + \frac{w_b \left(x_b - \mu_{\mathcal{A}}\right) - w_c \left(x_c - \mu_{\mathcal{A}}\right)}{W_{\mathcal{D}}}.
\end{aligned}$$

The update formula for $\mathcal{S}_{\mathcal{D},k}$ is

$$\begin{aligned}
\mathcal{S}_{\mathcal{D},k} &= \mathcal{S}_{\mathcal{A},k} + \mathcal{S}_{\mathcal{B},k} - \mathcal{S}_{\mathcal{C},k} + W_{\mathcal{A}} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^k + w_b \left(x_b - \mu_{\mathcal{D}}\right)^k - w_c \left(x_c - \mu_{\mathcal{D}}\right)^k \\
&\quad + \sum_{2 \le j < k} \binom{k}{j} \left\{\mathcal{S}_{\mathcal{A},j} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^{k-j} + \mathcal{S}_{\mathcal{B},j} \left(x_b - \mu_{\mathcal{D}}\right)^{k-j} - \mathcal{S}_{\mathcal{C},j} \left(x_c - \mu_{\mathcal{D}}\right)^{k-j}\right\}.
\end{aligned}$$

In the $k = 2$ case this simplifies to

$$\begin{aligned}
\mathcal{S}_{\mathcal{D},2} &= \mathcal{S}_{\mathcal{A},2} + W_{\mathcal{A}} \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^2 + w_b \left(x_b - \mu_{\mathcal{D}}\right)^2 - w_c \left(x_c - \mu_{\mathcal{D}}\right)^2, \\
&= \mathcal{S}_{\mathcal{A},2} + \left(\mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)^2 \left(W_{\mathcal{A}} - W_{\mathcal{D}}\right) + w_b \left(x_b - \mu_{\mathcal{D}}\right) \left(x_b - \mu_{\mathcal{A}}\right) - w_c \left(x_c - \mu_{\mathcal{D}}\right) \left(x_c - \mu_{\mathcal{A}}\right), \\
&= \mathcal{S}_{\mathcal{A},2} + \frac{W_{\mathcal{A}} w_b \left(x_b - \mu_{\mathcal{A}}\right)^2 - W_{\mathcal{D}} w_c \left(x_c - \mu_{\mathcal{D}}\right)^2}{W_{\mathcal{A}} + w_b}.
\end{aligned}$$

Further simplifications are possible when $w_b = w_c$ (as occurs in the vanilla case of equal weighted moments), and we arrive at

$$\mu_{\mathcal{D}} = \mu_{\mathcal{A}} + \frac{w_b\left(x_b - x_c\right)}{W_{\mathcal{D}}},$$
$$\mathcal{S}_{\mathcal{D},2} = \mathcal{S}_{\mathcal{A},2} + w_b\left(x_b + x_c - \mu_{\mathcal{A}} - \mu_{\mathcal{D}}\right)\left(x_b - x_c\right).$$

2 Vector Moments

Consider now the case where the data are a m dimensional vector, that is we observe $\mathbf{x}_1, \mathbf{x}_2$, and so on. Moments will be expressed in terms of the *Kronecker product*. While defined for general matrices, we use the Kronecker product only on vectors. Given vectors $\mathbf{a}$ and $\mathbf{b}$, we write

$$\mathbf{a} \otimes \mathbf{b} = \begin{bmatrix} a_1\mathbf{b} \\ a_2\mathbf{b} \\ a_3\mathbf{b} \\ \vdots \\ a_k\mathbf{b} \end{bmatrix}.$$

If $\mathbf{a}, \mathbf{b}$ are both m-dimensional then $\mathbf{a} \otimes \mathbf{b}$ is an m^2-dimensional vector.

Note that the Kronecker product is not commutative. However, there is a special matrix, the *Commutation matrix*, $\mathcal{K}_{m,m}$, that performs commutation [3]:

$$\mathbf{a} \otimes \mathbf{b} = \mathcal{K}_{m,m}\left(\mathbf{b} \otimes \mathbf{a}\right).$$

The subscript refers to the number of columns of the vectors in the Kronecker product, here m and m, but the matrix $\mathcal{K}_{m,m}$ is $m^2 \times m^2$. The Commutation matrix is involutary, that is $\mathcal{K}_{m,m}^2 = \mathbf{I}$.

We use special notation to denote the repeated application of Kronecker product to the same vector:

$$\mathbf{a}^{\otimes n} = \bigotimes_{i=1}^{n} \mathbf{a} = \underbrace{\mathbf{a} \otimes \mathbf{a} \otimes \ldots \mathbf{a}}_{n \text{ times}}.$$

If $\mathbf{a}$ is m-dimensional, then $\mathbf{a}^{\otimes n}$ is m^n-dimensional.

We can now quote a binomial expansion for Kronecker products.

$$\left(\mathbf{a} + \mathbf{b}\right)^{\otimes k} = \sum_{j}^{k} \mathbf{C}_{j,k}\mathbf{a}^{\otimes j} \otimes \mathbf{b}^{\otimes k-j},$$

for some binomial coefficient matrices, $\mathbf{C}_{j,k}$. These matrices are $m^k \times m^k$ matrices when $\mathbf{a}, \mathbf{b}$ are m-dimensional vectors. We have the following base and recurrence relationships between the coefficient matrices:

$$\mathbf{C}_{0,k} = \mathbf{C}_{k,k} = \mathbf{I}_{m^k},$$
$$\mathbf{C}_{j,k} = \mathbf{I}_m \otimes \mathbf{C}_{j-1,k-1} + \mathcal{K}_{m^{k-1},m}\left(\mathbf{C}_{j,k-1} \otimes \mathbf{I}_m\right), \quad \text{for } 0 < j < k.$$

Thus in particular,

$$\begin{aligned}
(\mathbf{a}+\mathbf{b})^{\otimes 2} &= \mathbf{I}_{m^2}\mathbf{a}^{\otimes 2} + \left(\mathbf{I}_{m^2} + \mathcal{K}_{m,m}\right)\mathbf{a}\otimes\mathbf{b} + \mathbf{I}_{m^2}\mathbf{b}^{\otimes 2},\\
(\mathbf{a}+\mathbf{b})^{\otimes 3} &= \mathbf{I}_{m^3}\mathbf{a}^{\otimes 3} + \left(\mathbf{I}_{m^3} + \mathcal{K}_{m^2,m} + \mathcal{K}_{m^2,m}\left(\mathcal{K}_{m,m}\otimes\mathbf{I}_m\right)\right)\mathbf{a}^{\otimes 2}\otimes\mathbf{b}\\
&\quad + \left(\mathbf{I}_{m^3} + \mathbf{I}_m\otimes\mathcal{K}_{m,m} + \mathcal{K}_{m^2,m}\right)\mathbf{a}\otimes\mathbf{b}^{\otimes 2} + \mathbf{I}_{m^3}\mathbf{b}^{\otimes 3},
\end{aligned}$$

and so on. Notice how the coefficent matrices generalize the scalar binomial coefficients. For example,

$$\mathbf{C}_{1,3}\mathbf{a}\otimes\mathbf{b}^{\otimes 2} = \mathbf{a}\otimes\mathbf{b}\otimes\mathbf{b} + \mathbf{b}\otimes\mathbf{a}\otimes\mathbf{b} + \mathbf{b}\otimes\mathbf{b}\otimes\mathbf{a}.$$

2.1 Update Formula for Vector Moments

We can now proceed as in Section 1. Let $\mathcal{A}$ be a set of indices over the data $\mathbf{x}_i$ and corresponding weights $w_i > 0$ Define the total elements, sum of weights, and (weighted) mean over $\mathcal{A}$ via

$$\begin{aligned}
n_{\mathcal{A}} &= |\mathcal{A}|,\\
W_{\mathcal{A}} &= \sum_{i\in\mathcal{A}} w_i,\\
\boldsymbol{\mu}_{\mathcal{A}} &= \frac{\sum_{i\in\mathcal{A}} w_i\mathbf{x}_i}{W_{\mathcal{A}}}.
\end{aligned}$$

Then go on to define the k^{th} centered weighted sum via

$$\mathcal{S}_{\mathcal{A},k} = \sum_{i\in\mathcal{A}} w_i(\mathbf{x}_i - \boldsymbol{\mu}_{\mathcal{A}})^{\otimes k}.$$

Again, $\mathcal{S}_{\mathcal{A},0} = W_{\mathcal{A}}$, and $\mathcal{S}_{\mathcal{A},1} = 0$.

Let $\mathcal{B}$ and $\mathcal{C}$ be sets of indices with the restriction that $\mathcal{A}\cap\mathcal{B} = \emptyset$, $\mathcal{C}\subseteq\mathcal{A}$, and define

$$\mathcal{D} = \mathcal{A}\cup\mathcal{B}\setminus\mathcal{C}.$$

We have:

$$\begin{aligned}
n_{\mathcal{D}} &= n_{\mathcal{A}} + n_{\mathcal{B}} - n_{\mathcal{C}},\\
W_{\mathcal{D}} &= W_{\mathcal{A}} + W_{\mathcal{B}} - W_{\mathcal{C}},\\
\boldsymbol{\mu}_{\mathcal{D}} &= \frac{W_{\mathcal{A}}\boldsymbol{\mu}_{\mathcal{A}} + W_{\mathcal{B}}\boldsymbol{\mu}_{\mathcal{B}} - W_{\mathcal{C}}\boldsymbol{\mu}_{\mathcal{C}}}{W_{\mathcal{D}}},\\
&= \boldsymbol{\mu}_{\mathcal{A}} + \frac{W_{\mathcal{B}}\left(\boldsymbol{\mu}_{\mathcal{B}} - \boldsymbol{\mu}_{\mathcal{A}}\right) - W_{\mathcal{C}}\left(\boldsymbol{\mu}_{\mathcal{C}} - \boldsymbol{\mu}_{\mathcal{A}}\right)}{W_{\mathcal{D}}},
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{S}_{\mathcal{D},k} &= \sum_{i\in\mathcal{D}} w_i(\mathbf{x}_i - \boldsymbol{\mu}_{\mathcal{D}})^{\otimes k},\\
&= \mathcal{S}_{\mathcal{A},k} + \mathcal{S}_{\mathcal{B},k} - \mathcal{S}_{\mathcal{C},k}\\
&\quad + W_{\mathcal{A}}(\boldsymbol{\mu}_{\mathcal{A}} - \boldsymbol{\mu}_{\mathcal{D}})^{\otimes k} + W_{\mathcal{B}}(\boldsymbol{\mu}_{\mathcal{B}} - \boldsymbol{\mu}_{\mathcal{D}})^{\otimes k} - W_{\mathcal{C}}(\boldsymbol{\mu}_{\mathcal{C}} - \boldsymbol{\mu}_{\mathcal{D}})^{\otimes k}\\
&\quad + \sum_{2\le j<k}\mathbf{C}_{j,k}\left\{\mathcal{S}_{\mathcal{A},j}\otimes\left(\boldsymbol{\mu}_{\mathcal{A}} - \boldsymbol{\mu}_{\mathcal{D}}\right)^{\otimes k-j} + \mathcal{S}_{\mathcal{B},j}\otimes\left(\boldsymbol{\mu}_{\mathcal{B}} - \boldsymbol{\mu}_{\mathcal{D}}\right)^{\otimes k-j}\right\}\\
&\quad - \sum_{2\le j<k}\mathbf{C}_{j,k}\left\{\mathcal{S}_{\mathcal{C},j}\otimes\left(\boldsymbol{\mu}_{\mathcal{C}} - \boldsymbol{\mu}_{\mathcal{D}}\right)^{\otimes k-j}\right\}.
\end{aligned}$$

This formula is a bit unweildy for the general case. Moreover, the number of elements in $\mathcal{S}_{\mathcal{D},k}$ quickly grows in k, even when accounting for symmetry.

The case $k = 2$ is of the highest interest. The matrix binomial coefficients do not enter into the computation, and we merely have:

$$\begin{aligned}\mathcal{S}_{\mathcal{D},2} &= \mathcal{S}_{\mathcal{A},2} + \mathcal{S}_{\mathcal{B},2} - \mathcal{S}_{\mathcal{C},2}\\ &\quad + W_{\mathcal{A}}(\boldsymbol{\mu}_{\mathcal{A}} - \boldsymbol{\mu}_{\mathcal{D}})^{\otimes 2} + W_{\mathcal{B}}(\boldsymbol{\mu}_{\mathcal{B}} - \boldsymbol{\mu}_{\mathcal{D}})^{\otimes 2} - W_{\mathcal{C}}(\boldsymbol{\mu}_{\mathcal{C}} - \boldsymbol{\mu}_{\mathcal{D}})^{\otimes 2}.\end{aligned}$$

2.2 Adding a single observation

We now consider the special case where $\mathcal{C}$ is empty, and $\mathcal{B}$ consists of the single index b. We then have

$$\begin{aligned}W_{\mathcal{D}} &= W_{\mathcal{A}} + w_b,\\ \boldsymbol{\mu}_{\mathcal{D}} &= \boldsymbol{\mu}_{\mathcal{A}} + w_b \frac{\mathbf{x}_b - \boldsymbol{\mu}_{\mathcal{A}}}{W_{\mathcal{D}}}.\end{aligned}$$

For the $k = 2$ case the centered moment is updated as

$$\begin{aligned}\mathcal{S}_{\mathcal{D},2} &= \mathcal{S}_{\mathcal{A},2} + \frac{W_{\mathcal{A}} w_b}{W_{\mathcal{D}}} (\mathbf{x}_b - \boldsymbol{\mu}_{\mathcal{A}})^{\otimes 2},\\ &= \mathcal{S}_{\mathcal{A},2} + W_{\mathcal{D}} \left(\mathbf{x}_b - \boldsymbol{\mu}_{\mathcal{D}}\right) \otimes \left(\boldsymbol{\mu}_{\mathcal{D}} - \boldsymbol{\mu}_{\mathcal{A}}\right).\end{aligned}$$

Note that the quantity $W_{\mathcal{D}} \left(\boldsymbol{\mu}_{\mathcal{D}} - \boldsymbol{\mu}_{\mathcal{A}}\right) = w_b \left(\mathbf{x}_b - \boldsymbol{\mu}_{\mathcal{A}}\right)$ is computed as an intermediary in updating the weighted mean, resulting in some computational savings.

2.3 Removing a single observation

As noted above, removing a single observation should look just like adding a single observation, except signs on the weights and weighted sums are flipped. Let c be the single index in $\mathcal{C}$. We then have $W_{\mathcal{D}} = W_{\mathcal{A}} - w_c$, and

$$\mu_{\mathcal{D}} = \boldsymbol{\mu}_{\mathcal{A}} - w_c \frac{\mathbf{x}_c - \boldsymbol{\mu}_{\mathcal{A}}}{W_{\mathcal{D}}}.$$

For the $k = 2$ case the centered sum computation simplifies to

$$\mathcal{S}_{\mathcal{D},2} = \mathcal{S}_{\mathcal{A},2} + W_{\mathcal{D}} \left(\mathbf{x}_c - \boldsymbol{\mu}_{\mathcal{D}}\right) \otimes \left(\boldsymbol{\mu}_{\mathcal{D}} - \boldsymbol{\mu}_{\mathcal{A}}\right).$$

2.4 Adding and removing a single observation

Consider the case of adding a single $b \in \mathcal{B}$ and removing a single $c \in \mathcal{C}$. We have

$$\begin{aligned}W_{\mathcal{D}} &= W_{\mathcal{A}} + w_b - w_c,\\ \boldsymbol{\mu}_{\mathcal{D}} &= \boldsymbol{\mu}_{\mathcal{A}} + \frac{w_b \left(\mathbf{x}_b - \boldsymbol{\mu}_{\mathcal{A}}\right) - w_c \left(\mathbf{x}_c - \boldsymbol{\mu}_{\mathcal{A}}\right)}{W_{\mathcal{D}}}.\end{aligned}$$

The update formula for $\mathcal{S}_{\mathcal{D},2}$ is

$$\mathcal{S}_{\mathcal{D},2} = \mathcal{S}_{\mathcal{A},2} + \frac{W_{\mathcal{A}} w_b (\mathbf{x}_b - \boldsymbol{\mu}_{\mathcal{A}})^{\otimes 2} - W_{\mathcal{D}} w_c (\mathbf{x}_c - \boldsymbol{\mu}_{\mathcal{D}})^{\otimes 2}}{W_{\mathcal{A}} + w_b}.$$

Because Kronecer products are not commutative, the trick to simplify this expression in the case where $w_b = w_c$, involving completing the square, cannot be easily applied.

3 The output

Several kinds of output are produced by the fromo package. We describe them in more detail here.

mean The (weighted) mean over the index set $\mathcal{A}$ is simply $\mu_{\mathcal{A}}$.

standard deviation The standard deviation over the index set $\mathcal{A}$ is

$$\sigma_{\mathcal{A}} = \sqrt{\frac{\mathcal{S}_{\mathcal{A},2}}{W_{\mathcal{A}} - \nu}},$$

where ν are the 'consumed degrees of freedom', and is typically set to 1. When the normalization flag is set to true, however, it is intended that the weights be normalized to have mean value 1. In this case the standard deviation over $\mathcal{A}$ is defined as

$$\sigma_{\mathcal{A}} = \sqrt{\frac{\mathcal{S}_{\mathcal{A},2}}{W_{\mathcal{A}}} \frac{n_{\mathcal{A}}}{n_{\mathcal{A}} - \nu}}.$$

When $\nu = 0$, these are identical.

centered moment The k^{th} centered moment is defined as

$$\mathcal{M}_{\mathcal{A},k} = \frac{\mathcal{S}_{\mathcal{A},k}}{W_{\mathcal{A}}}.$$

Note that for $k > 2$ we do not support consumed degrees of freedom, and so normalization of weights has no effect. The first centered moment is zero: $\mathcal{M}_{\mathcal{A},1} = 0$.

standardized moment The k^{th} standardized (and centered) moment is defined as

$$\mathcal{Y}_{\mathcal{A},k} = \frac{\mathcal{M}_{\mathcal{A},k}}{\sigma_{\mathcal{A}}^k}.$$

Normalization and consumed degrees of freedom affect the computation of $\sigma_{\mathcal{A}}^k$, and so affect the standardized moments.

centered value For a given index i and some fixed set $\mathcal{A}$, the centered version of x_i is merely $x_i - \mu_{\mathcal{A}}$.

standardized value For a given index i and some fixed set $\mathcal{A}$, the standardized version of x_i is merely $x_i/\sigma_{\mathcal{A}}$. It is affected by normalization and consumed degrees of freedom.

z-scored value For a given index i and some fixed set $\mathcal{A}$, the z-scored version of x_i is $(x_i - \mu_{\mathcal{A}})/\sigma_{\mathcal{A}}$. It is affected by normalization and consumed degrees of freedom.

Cumulants Sometimes called "centered cumulants" in the package, though this is redundant and not common usage. Cumulants are defined from the centered moments by the recursive formula: [5]

$$\mathcal{K}_{\mathcal{A},r} = \mathcal{M}_{\mathcal{A},r} - \sum_{j=1}^{r-2} \binom{r-1}{j} \mathcal{M}_{\mathcal{A},j} \mathcal{K}_{\mathcal{A},r-j}. \tag{1}$$

Note that the sum is over an empty index for $r = 2$, and that $\mathcal{M}_{\mathcal{A},1} = 0$, so we have

$$\begin{aligned}
\mathcal{K}_{\mathcal{A},2} &= \mathcal{M}_{\mathcal{A},2}, \\
\mathcal{K}_{\mathcal{A},3} &= \mathcal{M}_{\mathcal{A},3}, \\
\mathcal{K}_{\mathcal{A},4} &= \mathcal{M}_{\mathcal{A},4} - 3{\mathcal{M}_{\mathcal{A},2}}^2, \\
\mathcal{K}_{\mathcal{A},5} &= \mathcal{M}_{\mathcal{A},5} - 10\mathcal{M}_{\mathcal{A},3}\mathcal{M}_{\mathcal{A},2},
\end{aligned}$$

and so on.

Standardized Cumulants These are the regular cumulants normalized by the computed standard deviation:

$$\mathcal{G}_{\mathcal{A},r} = \frac{\mathcal{K}_{\mathcal{A},r}}{\sigma_{\mathcal{A}}^r}. \tag{2}$$

3.1 Output for Bivariate Input

Consider now the case where the user provides input x_i and y_i. We denote the (weighted) mean over the index set $\mathcal{A}$ of the x_i as $\mu_{\mathcal{A},x}$, and similarly $\mu_{\mathcal{A},y}$ is the (weighted) mean of the y_i. Write $\mathcal{S}_{\mathcal{A},x,x}$, $\mathcal{S}_{\mathcal{A},x,y}$, and $\mathcal{S}_{\mathcal{A},y,y}$ for the weighted sums of centered second power terms:

$$\begin{aligned}
\mathcal{S}_{\mathcal{A},x,x} &= \sum_{i\in\mathcal{A}} w_i \left(x_i - \mu_{\mathcal{A},x}\right)^2, \\
\mathcal{S}_{\mathcal{A},x,y} &= \sum_{i\in\mathcal{A}} w_i \left(x_i - \mu_{\mathcal{A},x}\right)\left(y_i - \mu_{\mathcal{A},y}\right), \\
\mathcal{S}_{\mathcal{A},y,y} &= \sum_{i\in\mathcal{A}} w_i \left(y_i - \mu_{\mathcal{A},y}\right)^2.
\end{aligned}$$

We then have the following output:

correlation The correlation is computed as

$$\frac{\mathcal{S}_{\mathcal{A},x,y}}{\sqrt{\mathcal{S}_{\mathcal{A},x,x}\mathcal{S}_{\mathcal{A},y,y}}}.$$

regression slope We consider the ordinary least squares regression of y_i against x_i. The data are supposed to follow $y_i = \beta_0 + \beta_1 x_i$, and this computation estimates β_1. This has the value

$$\frac{\mathcal{S}_{\mathcal{A},x,y}}{\mathcal{S}_{\mathcal{A},x,x}}.$$

regression intercept We again consider the OLS regression of y_i against x_i and return an estimate of β_0 taking value

$$\mu_{\mathcal{A},y} - \mu_{\mathcal{A},x} \frac{\mathcal{S}_{\mathcal{A},x,y}}{\mathcal{S}_{\mathcal{A},x,x}}.$$

regression standard error We again consider the OLS regression of y_i against x_i. We write σ^2 to denote the variance of $y_i - (\beta_0 + \beta_1 y_i)$. The "regression standard error" is an estimate of σ. It is computed as

$$\sqrt{\frac{\mathcal{S}_{\mathcal{A},y,y} - \mathcal{S}^2_{\mathcal{A},x,y}/\mathcal{S}_{\mathcal{A},x,x}}{n_{\mathcal{A}} - \nu}},$$

where ν are the 'consumed degrees of freedom'. When $\nu = 0$ this quantity is often denoted as $\hat{\sigma}^2$; when $\nu = 2$ it is often written as s^2.

regression slope standard error This estimates the standard error of the estimate of β_1 in the regression. It is computed as

$$\sqrt{\frac{s^2}{\mathcal{S}_{\mathcal{A},x,x}}},$$

where s^2 is the regression standard error computed with $\nu = 2$.

regression intercept std. err. This estimates the standard error of the estimate of β_0 in the regression. It is computed as

$$\sqrt{\frac{s^2 \left(\mathcal{S}_{\mathcal{A},x,x}/n_{\mathcal{A}} + \mu^2_{\mathcal{A},x}\right)}{\mathcal{S}_{\mathcal{A},x,x}}},$$

where s^2 is the regression standard error computed with $\nu = 2$.

covariance We estimate the covariance of y_i and x_i by computing

$$\frac{\mathcal{S}_{\mathcal{A},x,y}}{n_{\mathcal{A}} - \nu},$$

where ν are the 'consumed degrees of freedom', typically set to 2.

covariance 3 We estimate the full covariance matrix of y_i and x_i by computing the lower triangle of the matrix

$$\frac{1}{n_{\mathcal{A}} - \nu} \left[\begin{array}{cc} \mathcal{S}_{\mathcal{A},x,x} & \mathcal{S}_{\mathcal{A},x,y} \\ \mathcal{S}_{\mathcal{A},x,y} & \mathcal{S}_{\mathcal{A},y,y} \end{array} \right],$$

where ν are the 'consumed degrees of freedom', typically set to 2.

4 Running Moments

The fromo package can also perform running[1] computations of moments, centered moments, centered values, standardized values, z-scored values and so on.

[1] Variously known also as *rolling*, *boxcar*, or *finite impulse response* computations.

Given an integral window, W, and data and weights vectors x_i, w_i, we compute output y_i as follows: let $\mathcal{D}$ be the set of j with $i - W < j \le i$. Then compute the desired moment over $\mathcal{D}$, and return it as y_i. The 'comparison' operations (computed centered, standardized, or z-scored versions of a variable) admit a *lookahead* parameter, l. In this case, we let $\mathcal{D}$ be the set of j with $i - W + l < j \le i + l$, then compute the moments over $\mathcal{D}$ and normalize x_i with respect to those moments. When $l > 0$, we are using future information to compute y_i. That is, y_i will depend on x_j with $j > i$. This is not the case for the typical moments computations or for the case where the lookahead is non-positive.

We also support a time (or other counter) based running computation. Here the input are the data, and weights vectors, x_i, w_i, but also a vector of time indices, say t_i which are non-decreasing: $t_1 \le t_2 \le \ldots$. It is assumed that t_0 is essentially $-\infty$. The window, W is now a time-based window. The output y_i is defined in terms of the moments computations over the set $\mathcal{D}$ which are all j such that $t_i - W < t_j \le t_i$. For comparison functions, we again admit a lookahead so that $\mathcal{D}$ are all j such that $t_i - W + l < t_j \le t_i + l$. Obviously if we let $t_i = i$, we recover the 'vanilla' running moments computations. We also support supplying the t_i implicitly as the sum of positive deltas: let vector d_i of strictly positive elements be given. Then we assume

$$t_i = \sum_{1 \le j \le i} d_j,$$

and then perform the time-based running computation. We also support the case where the d_i are actually just the w_i.

References

[1] J. Bennett, R. Grout, P. Pebay, D. Roe, and D. Thompson. Numerically stable, single-pass, parallel statistics algorithms. In *2009 IEEE International Conference on Cluster Computing and Workshops*, pages 1–8, Aug 2009. doi: 10.1109/CLUSTR.2009.5289161.

[2] John D. Cook. Accurately computing running variance. URL http://www.johndcook.com/standard_deviation.html.

[3] Jan R. Magnus and H. Neudecker. The commutation matrix: Some properties and applications. *Ann. Statist.*, 7(2):381–394, 03 1979. doi: 10.1214/aos/1176344621. URL https://doi.org/10.1214/aos/1176344621.

[4] Timothy B. Terriberry. Computing higher-order moments online, 2008. URL http://people.xiph.org/~tterribe/notes/homs.html.

[5] R. Willink. Relationships between central moments and cumulants, with formulæ for the central moments of Gamma distributions. *Communications in Statistics - Theory and Methods*, 32(4):701–704, 2003. doi: https://doi.org/10.1081/STA-120018823. URL https://doi.org/10.1081/STA-120018823.