Moving-Window-Statistics

Atharva Aalok

30 August 2023

1 Formulas

Consider a time series $X = [x_1, x_2, \; ..., x_i, \; ..., x_j, \; ..., x_N]$. Also, consider a moving window over the data of size *window_size* which starts at index $i+1$ (that is i is the index just before the window) and ends at index j. Therefore, $j - i = window\_size$. Also, let's denote the size of the step by which the window moves over the data to be *window_step*. That is, the next window starts at $i + 1 + window\_step$ and ends at $j + window\_step$.

A simple and naive approach towards calculating moving window statistics could be to move the window over the data and for each window calculate the moving window statistic using in-built functions (such as var, skewness etc. in MATLAB). But this approach is very slow as it involves a lot of redundant calculations. For example, consider the case of finding the moving variance. Using this naive approach, for each window one would have to square all the data points towards calculation of the variance. Now since the windows will have a good amount of overlap the same data point which is common amongst several windows will be squared repeatedly, leading to redundant calculations which slow down the process.

To calculate the different moving window statistics fast we use a one-pass approach where we go through each data point only once, getting rid of all redundant operations. The statistics can be calculated using the below given formulas, the proofs for which are provided later in the Proofs section.

Define S_k^j to be the sum of the k^{th} powers of the first j x_n's. In particular, we will require the following:

$$S_1^j = \sum_{n=1}^{j} x_n$$

$$S_2^j = \sum_{n=1}^{j} x_n^2$$

$$S_3^j = \sum_{n=1}^{j} x_n^3$$

$$S_4^j = \sum_{n=1}^{j} x_n^4$$

The central moments for the data points within a window can be calculated using the following formulas:

$$M_1^{j-i} = \frac{S_1^j - S_1^i}{j-i} = \frac{jM_1^j - iM_1^i}{j-i}$$

$$M_2^{j-i} = (S_2^j - S_2^i) - (j-i)(M_1^{j-i})^2$$

$$M_3^{j-i} = (S_3^j - S_3^i) - 3M_1^{j-i}(S_2^j - S_2^i) + 2(j-i)(M_1^{j-i})^3$$

$$M_4^{j-i} = (S_4^j - S_4^i) - 4M_1^{j-i}(S_3^j - S_3^i) + 6(M_1^{j-i})^2(S_2^j - S_2^i) - 3(j-i)(M_1^{j-i})^4$$

For calculating autocorrelation at some given lag l, we will additionally require the product of the data with a delayed copy of itself. Thus the following definition will be useful:

$$P_l^j = \sum_{n=1}^{j-l} x_n x_{n+l}$$

Using the above defined quantities, the moving window statistical quantities can be then computed using the below provided formulas. Here we use N in place of $window\_size$, therefore $N = widnow\_size = j - i$.

$$\text{Mean} = M_1^j$$
$$\text{Moving mean} = M_1^{j-i}$$

$$\text{RMS} = \sqrt{\frac{S_2^j}{N}}$$
$$\text{Moving RMS} = \sqrt{\frac{S_2^j - S_2^i}{N}}$$

$$\text{Variance} = \frac{M_2^j}{N-1}$$
$$\text{Moving variance} = \frac{M_2^{j-i}}{N-1}$$

$$\text{Skewness} = \frac{\sqrt{N} M_3^j}{(M_2^j)^{\frac{3}{2}}}$$
$$\text{Moving skewness} = \frac{\sqrt{N} M_3^{j-i}}{(M_2^{j-i})^{\frac{3}{2}}}$$

$$\text{Kurtosis} = \frac{N M_4^j}{(M_2^j)^2}$$
$$\text{Moving kurtosis} = \frac{N M_4^{j-i}}{(M_2^{j-i})^2}$$

$$\text{Autocorrelation} = \frac{P_l^j - M_1^j(S_1^j + S_1^{j-l} - S_1^l) + (M_1^j)^2(j-l)}{M_2^j}$$
$$\text{Moving autocorrelation} = \frac{P_l^{j-1} - P_l^i - M_1^{j-i}(S_1^j + S_1^{j-l} - S_1^{i+l} - S_1^i) + (M_1^{j-i})^2(N-l)}{M_2^{j-i}}$$

Note: The usage of N or $N-1$ are such that the statistic values match the values obtained from the built-in functions of MATLAB.

2 Proofs

Consider a time series $X = [x_1, x_2, \; ..., x_i, \; ..., x_j, \; ..., x_N]$. We start with some necessary definitions and notations that will help us in calculating the moving window statistics.

As previously, we use S_k^i to denote the sum of k^{th} powers of first i points. That is,

$$S_k^i = \sum_{n=1}^{i} x_n^k$$

Let's use the notation M_1^i to denote the mean of the first i points. That is,

$$M_1^i = \frac{x_1 + x_2 + ... + x_i}{i} = \frac{\sum_{n=1}^{i} x_n}{i}$$

Also, with a slight difference let's use the notation M_k^i to denote the k^{th} central moment of the first i data points. That is,

$$M_k^i = \sum_{n=1}^{i}(x_n - \bar{x})^k = \sum_{n=1}^{i}(x_n - M_1^i)^k$$

Note that M_1^i is different from the other M_k^i in definition in that it is scaled down by i and neither is it central (the central average would be just 0). Though it might be slightly confusing, this notation is quite intuitive and will be helpful later.

Now we define notation for window quantities. Consider that i is the index just before the window (that is, window begins at $i+1$) and j is the last index of the window. Then $j - i = window\_size$.

We denote the equivalent of M_1^i for a window spanning $i+1$ to j as:

$$M_1^{j-i} = \frac{x_{i+1} + ... + x_j}{j - i}$$

This is the mean of the data points within the window. The $j-i$ in the super script is to denote the data between the indices i and j. We can now use this to define the higher central moments for the window.

Define,

$$M_k^{j-i} = \sum_{n=i+1}^{j}(x_n - M_1^{j-i})^k$$

The M_1^{j-i} appears as it is the mean of the data points from $i+1$ to j. The M_k^{j-i} are useful for calculating our moving window statistics, as shown before in the Formulas section.

We now derive the formulas for evaluating these M_k^{j-i}'s. The formula for M_1^{j-i} is simply:

$$\boxed{M_1^{j-i} = \frac{S_1^j - S_1^i}{j-i} = \frac{jM_1^j - iM_1^i}{j-i}} \tag{1}$$

The higher central moments for the windows will now utilize M_1^{j-i} just as the central moments of first i data, M_k^i utilized M_1^i. Let's define,

$$M_2^{j-i} = \sum_{n=i+1}^{j}(x_n - M_1^{j-i})^2$$

M_1^{j-i} is constant for the window from $i+1$ to j, hence we can take it out of the summation.

$$\begin{aligned}
M_2^{j-i} &= \sum_{n=i+1}^{j} x_n^2 - 2M_1^{j-i}\sum_{n=i+1}^{j} x_n + (M_1^{j-i})^2\sum_{n=i+1}^{j} 1 \\
&= \sum_{n=i+1}^{j} x_n^2 - 2M_1^{j-i}(j-i)M_1^{j-i} + M_1^{j-i}(j-i)M_1^{j-i} \\
&= \sum_{n=i+1}^{j} x_n^2 - (j-i)M_1^{j-i}
\end{aligned}$$

We can rewrite the summation of x_n^2's here using S_2^n's as follows:

$$\boxed{\begin{aligned}
M_2^{j-i} &= \sum_{n=i+1}^{j} x_n^2 - (j-i)M_1^{j-i} & (2) \\
&= (S_2^j - S_2^i) - (j-i)(M_1^{j-i})^2 & (3)
\end{aligned}}$$

The formula for M_3^{j-i} can be similarly found as:

$$
\begin{aligned}
M_3^{j-i} &= \sum_{n=i+1}^{j} (x_n - M_1^{j-i})^3 \\
&= \sum_{n=i+1}^{j} x_n^3 - 3M_1^{j-i} \sum_{n=i+1}^{j} x_n^2 + 3(M_1^{j-i})^2 \sum_{n=i+1}^{j} x_n - (M_1^{j-i})^3 \sum_{n=i+1}^{j} 1 \\
&= \sum_{n=i+1}^{j} x_n^3 - 3M_1^{j-i} \sum_{n=i+1}^{j} x_n^2 + 3(M_1^{j-i})^2 (j-i) M_1^{j-i} - (M_1^{j-i})^3 (j-i) \\
&= \sum_{n=i+1}^{j} x_n^3 - 3M_1^{j-i} \sum_{n=i+1}^{j} x_n^2 + 2(j-i)(M_1^{j-i})^3
\end{aligned}
$$

We can rewrite the summations of powers of x_n's using corresponding S_k^n's as:

$$
\boxed{
\begin{aligned}
M_3^{j-i} &= \sum_{n=i+1}^{j} x_n^3 - 3M_1^{j-i} \sum_{n=i+1}^{j} x_n^2 + 2(j-i)(M_1^{j-i})^3 && (4) \\
&= (S_3^j - S_3^i) - 3M_1^{j-i}(S_2^j - S_2^i) + 2(j-i)(M_1^{j-i})^3 && (5)
\end{aligned}
}
$$

Finally, we can similarly derive the formula for M_4^{j-i} as follows:

$$
\begin{aligned}
M_4^{j-i} &= \sum_{n=i+1}^{j} (x_n - M_1^{j-i})^4 \\
&= \sum_{n=i+1}^{j} x_n^4 - 4M_1^{j-i} \sum_{n=i+1}^{j} x_n^3 + 6(M_1^{j-i})^2 \sum_{n=i+1}^{j} x_n^2 - 4(M_1^{j-i})^3 \sum_{n=i+1}^{j} x_n + (M_1^{j-i})^4 \sum_{n=i+1}^{j} 1 \\
&= \sum_{n=i+1}^{j} x_n^4 - 4M_1^{j-i} \sum_{n=i+1}^{j} x_n^3 + 6(M_1^{j-i})^2 \sum_{n=i+1}^{j} x_n^2 - 4(M_1^{j-i})^3 \sum_{n=i+1}^{j} (j-i)M_1^{j-i} + (M_1^{j-i})^4 (j-i) \\
&= \sum_{n=i+1}^{j} x_n^4 - 4M_1^{j-i} \sum_{n=i+1}^{j} x_n^3 + 6(M_1^{j-i})^2 \sum_{n=i+1}^{j} x_n^2 - 3(j-i)(M_1^{j-i})^4
\end{aligned}
$$

Rewriting using S_k^n's:

$$
\boxed{
\begin{aligned}
M_4^{j-i} &= \sum_{n=i+1}^{j} x_n^4 - 4M_1^{j-i} \sum_{n=i+1}^{j} x_n^3 + 6(M_1^{j-i})^2 \sum_{n=i+1}^{j} x_n^2 - 3(j-i)(M_1^{j-i})^4 && (6) \\
&= (S_4^j - S_4^i) - 4M_1^{j-i}(S_3^j - S_3^i) + 6(M_1^{j-i})^2(S_2^j - S_2^i) - 3(j-i)(M_1^{j-i})^4 && (7)
\end{aligned}
}
$$

These central moments for the window data points can now be used to calculate the moving window statistics using the expressions given in the Formulas section.

Now, we describe the method to calculate the autocorrelation of a time series given a particular value of lag. Towards this, we need to first clearly define how exactly the inbuilt *autocorr* function in MATLAB calculates it. Given the timeseries X as previously, it calculates autocorrelation at lag l as:

$$
AC_l(X) = \frac{1}{var(X)(N-1)} \sum_{i=1}^{N-l} (x_n - \bar{x})(x_{n+l} - \bar{x})
$$

Since, $\bar{x}$ for a window is just M_1^{j-i} we can rewrite the above as:

$$
AC_l(X) = \frac{1}{var(X)(N-1)} \sum_{i=1}^{N-l} (x_n - M_1^{j-i})(x_{n+l} - M_1^{j-i})
$$

Expanding this results in:

$$\begin{aligned}
AC_l(X) &= \frac{1}{var(X)(N-1)} \sum_{n=1}^{N-l} (x_n - M_1^{j-i})(x_{n+l} - M_1^{j-i}) \\
&= \frac{1}{var(X)(N-1)} \left\{ \sum_{n=1}^{N-l} x_n x_{n+l} - M_1^{j-i} \sum_{n=1}^{N-l} x_n - M_1^{j-i} \sum_{n=1}^{N-l} x_{n+l} + (M_1^{j-i})^2 \sum_{n=1}^{N-l} 1 \right\} \\
&= \frac{1}{var(X)(N-1)} \left\{ \sum_{n=1}^{N-l} x_n x_{n+l} - M_1^{j-i}(S_1^{j-l} - S_1^i) - M_1^{j-i}(S_1^j - S_1^i - S_1^{i+l} + S_1^i) + (N-l)(M_1^{j-i})^2 \right\} \\
&= \frac{1}{M_2^{j-i}(N-1)} \left\{ P_l^{j-1} - P_l^i - M_1^{j-i}(S_1^j + S_1^{j-l} - S_1^{i+l} - S_1^i) + (N-l)(M_1^{j-i})^2 \right\}
\end{aligned}$$

This can now be written using P_l^j as:

$$\boxed{AC_l(X) = \frac{1}{M_2^{j-i}(N-1)} \left\{ P_l^{j-1} - P_l^i - M_1^{j-i}(S_1^j + S_1^{j-l} - S_1^{i+l} - S_1^i) + (N-l)(M_1^{j-i})^2 \right\}}$$

(8)

(9)