Iterative Sample Statistic

Edoardo Pasca

October 25, 2016

Some time ago I was sick with some fever. The only thing I could do was to lay in bed and measure my body temperature. The thermometer took so little time to measure that I tried again and again. Each time the temperature was different.

I thought to measure a few times and then to calculate the average, to get a more stable result.

Doing it without a piece of paper starts to be a bit difficult when you have to deal with more than 5 samples (at least for me). I started thinking that there must be a way to update the last calculated average with the next sample.

1 Mean

The average is defined as:

$$\langle v \rangle_N = \frac{1}{N} \sum_{i=1}^{N} v_i \tag{1}$$

Adding one sample v_{N+1}, the next evaluation of the average is:

$$\begin{aligned} \bar{v}_{N+1} &= \frac{1}{N+1} \sum_{i=1}^{N+1} v_i \\ &= \frac{1}{N+1} \left(\sum_{i=1}^{N} v_i + v_{N+1} \right) \\ &= \frac{1}{N+1} \left(N \bar{v}_N + v_{N+1} \right) \\ &= \frac{N}{N+1} \bar{v}_N + \frac{v_{N+1}}{N+1} \end{aligned} \tag{2}$$

where $\bar{v}_{N+1}$ is the average calculated with $N+1$ samples and N is the number of samples. This formula says that we can calculate the next average by keeping in mind only 3 numbers: the previous average, the number of samplings and the last sample. That's nice!

2 Standard Deviation

The variance is defined as the square of the standard deviation σ:

$$\sigma_N^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(v_i - \bar{v}_N \right)^2 \tag{3}$$

Again, adding the $N+1$-th sample the estimation of the variance can be expressed as:

$$\sigma^2_{N+1} = \frac{1}{N-1} \sum_{i=1}^{N+1} (v_i - \bar{v}_{N+1})^2 \tag{4}$$

we now use the eq. 3 to express $\bar{v}_{N+1}$ in terms of $\bar{v}_N$

$$\sigma^2_{N+1} = \frac{1}{N} \sum_{i=1}^{N+1} \left(v_i - \left(\frac{N}{N+1}\bar{v}_N + \frac{v_{N+1}}{N+1} \right) \right)^2 \tag{5}$$

Before going further, we better arrange the term in the sum with a nice trick:

$$\begin{aligned} v_i - \left(\frac{N}{N+1}\bar{v}_N + \frac{v_{N+1}}{N+1} \right) = & \\ &= v_i - \left(\bar{v}_N - \bar{v}_N + \frac{N}{N+1}\bar{v}_N + \frac{v_{N+1}}{N+1} \right) \\ &= v_i - \bar{v}_N - \left(-\bar{v}_N + \frac{N}{N+1}\bar{v}_N + \frac{v_{N+1}}{N+1} \right) \\ &= (v_i - \bar{v}_N) - \left(\frac{N-(N+1)}{N+1}\bar{v}_N + \frac{v_{N+1}}{N+1} \right) \\ &= (v_i - \bar{v}_N) - \left(-\frac{1}{N+1}\bar{v}_N + \frac{v_{N+1}}{N+1} \right) \end{aligned} \tag{6}$$

The term in the sum is actually expressed by the sum of two terms, the first depending on the various samplings, v_i and the previous evaluation of the average, $\bar{v}_N$, and the second term depending only on the previous evaluation of the average, $\bar{v}_N$, the number of samplings, N, and the last sample, v_{N+1}.

$$a_i = v_i - \bar{v}_N \tag{7}$$

$$b_{N,\bar{v}_N} = b = \frac{v_{N+1} - \bar{v}_N}{N+1} \tag{8}$$

Now we can express the variance, eq. 5, with the help of these variables a_i and b as,

$$\sigma^2_{N+1} = \frac{1}{N} \left[\left(\sum_{i=1}^{N} (a_i - b)^2 \right) + (a_{N+1} - b)^2 \right] \tag{9}$$

$$\begin{aligned} &= \frac{1}{N} \left[\left(\sum_{i=1}^{N} \left(a_i^2 + b^2 - 2a_i b\right) \right) + (a_{N+1} - b)^2 \right] \\ &= \frac{1}{N} \left[\left(\sum_{i=1}^{N} \left(a_i^2 + b^2\right) \right) + (a_{N+1} - b)^2 \right] = \frac{1}{N} \left[\left(\sum_{i=1}^{N} a_i^2 \right) + Nb^2 + (a_{N+1} - b)^2 \right] \\ &= \frac{1}{N} \left[(N-1)\sigma^2_N + Nb^2 + (a_{N+1} - b)^2 \right] \\ &= \frac{1}{N} \left[(N-1)\sigma^2_N + \frac{N}{N+1} (v_{N+1} - \bar{v}_N)^2 \right] \end{aligned}$$

$$\sigma^2_{N+1} = \frac{N-1}{N}\sigma^2_N + \frac{1}{N+1} (v_{N+1} - \bar{v}_N)^2 \tag{10}$$

Eqs. (2) and (10) allow to iterative calculate the next estimation of average and variance using only the previous estimation, the last sampling and the number of samples.

3 Higher order moments

As last algebraic remark,

$$\begin{aligned} a_{N+1} - b =&(v_{N+1} - \bar{v}_N) - \frac{1}{N+1}(v_{N+1} - \bar{v}_N) \\ =&\frac{N}{N+1}(v_{N+1} - \bar{v}_N) \\ a_{N+1} - b =&Nb \end{aligned} \tag{11}$$

We can easily extend this treatment to higher standardized moments of the distribution of the population:

$$x_k = \frac{\sum_{i=1}^{N} (v_i - \bar{v}_N)^k}{\sigma_N^k} = \frac{M_{N,k}}{\sigma_N^k} \tag{12}$$

where k is the moment's order. It is now very easy to calculate the

$$\begin{aligned} M_{N+1,k} &= \sum_i^{N+1} (v_i - \bar{v}_{N+1})^k \\ &= \left[\sum_i^N (a_i - b)^k \right] + (a_{N+1} - b)^k \\ &= \left[\sum_i^N (a_i - b)^k \right] + (Nb)^k \end{aligned} \tag{13}$$

3.1 Skewness

The skewness is a third order moment, $k = 3$.

$$
\begin{aligned}
M_{N+1,3} &= \left[\sum_{i}^{N} (a_i - b)^3 \right] + (Nb)^3 \\
&= \left[\sum_{i}^{N} (a_i^3 - b^3 + 3a_i b^2 - 3a_i^2 b) \right] + (Nb)^3 \\
&= \left[\sum_{i}^{N} (a_i^3 - 3a_i^2 b) \right] - Nb^3 + (Nb)^3 \\
&= \left[\sum_{i}^{N} (a_i^3 - 3a_i^2 b) \right] + Nb^3(N^2 - 1) \\
&= M_{N,3} - 3b \left[\sum_{i}^{N} a_i^2 \right] + Nb^3(N^2 - 1) \\
&= M_{N,3} - 3b(N-1)\sigma_N^2 + Nb^3(N-1)(N+1)
\end{aligned}
\tag{14}
$$

3.2 Kurtosis

The kurtosis is a fourth order moment, $k = 4$.

$$
\begin{aligned}
M_{N+1,4} &= \left[\sum_{i}^{N} (a_i - b)^4 \right] + (Nb)^4 \\
&= \left[\sum_{i}^{N} (a_i^4 + b^4 - 4a_i^3 + 5a_i^2) \right] + (Nb)^4 \\
&= \left[\sum_{i}^{N} (a_i^3 - 3a_i^2 b) \right] - Nb^3 + (Nb)^4 \\
&= \left[\sum_{i}^{N} (a_i^3 - 3a_i^2 b) \right] + Nb^3(N^2 - 1) \\
&= M_{N,3} - 3b \left[\sum_{i}^{N} a_i^2 \right] + Nb^3(N^2 - 1) \\
&= M_{N,3} - 3b(N-1)\sigma_N^2 + Nb^3(N-1)(N+1)
\end{aligned}
\tag{15}
$$