# Recursion Formulas for Statistics — Sample Mean and Sample Variance

Paolo Bosetti

January 24, 2017

**Abstract**

This memo reports the calculation scheme that can be adopted for calculating sample mean and variance statistics by using a pair of recurrence formulas. This approach comes handy whenever you have to perform a continuous, inline assessment of those indicators with minimum memory footprint (e.g. on microcontrollers), without the need of storing the whole set of sample values.

## 1 Sample Mean

This is an easy one: given the stochastic variable $x$ and by indicating its sample mean for a set of $i$ observations as $\bar{x}_i$, we have:

$$\bar{x}_1 := x_1 \tag{1}$$

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

$$= \frac{n-1}{n}\left(\sum_{i=1}^{n-1}\frac{x_i}{n-1}\right) + \frac{x_n}{n} \tag{3}$$

$$= \frac{1}{n}\left((n-1)\bar{x}_{n-1} + x_n\right) \tag{4}$$

where $x_n$ is the current observation (after $n$ events), and $x_1$ is the very first observation.

By using Eq. 4, the sample mean value $\bar{x}_n$ can be continuously updated at every acquisition using only the last observation $x_n$, the previous value of the sample mean $\bar{x}_{n-1}$, and the total number of observations $n$. There is *no need for storing the whole set of observations*, and the algorithm complexity is $O(1)$.

## 2 Sample Variance

The recurrence formula for sample variance is a little more complex, and care must be payed in the formulation in order to avoid differences between small quantities, which may bring to large rounding errors.

By definition of sample variance for $n$ observations, $s_n^2$:

$$s_n^2 = \sum_{i=1}^{n} \frac{(\bar{x}_n - x_i)^2}{n-1} = \frac{SS_n}{n-1} \tag{5}$$

$$s_n = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right)} \tag{6}$$

where $SS_n$ is the *sum of squares* of the $n$ observations, which is the product of the sample variance times the number of degrees of freedom.

The sum of squares (which is the only part in the definition of sample variance that is depending on previous values) can be more conveniently expressed as:

$$SS_n = \left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right) \tag{7}$$

so that the increment in sum of squares can be obtained:

$$SS_n - SS_{n-1} = \sum_{1=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{1=1}^{n} x_i\right)^2 - \sum_{1=1}^{n-1} x_i^2 + \frac{1}{n-1}\left(\sum_{1=1}^{n-1} x_i\right)^2 \tag{8}$$

where we can substitute:

$$\sum_{1=1}^{n} x_i = n\bar{x}_n \tag{9}$$

$$\sum_{1=1}^{n-1} x_i = \sum_{1=1}^{n} x_i - x_n = n\bar{x}_n - x_n \tag{10}$$

$$\sum_{1=1}^{n} x_i^2 - \sum_{1=1}^{n-1} x_i^2 = x_n^2 \tag{11}$$

thus having:

$$SS_n - SS_{n-1} = x_n^2 - \frac{1}{n}(n\bar{x}_n)^2 + \frac{1}{n-1}(n\bar{x}_n - x_n)^2 \tag{12}$$

$$= x_n^2 - n\bar{x}_n^2 + \frac{1}{n-1}(n^2\bar{x}_n^2 - 2n\bar{x}_n x_n + x_n^2) \tag{13}$$

$$= \frac{1}{n-1}(nx_n^2 + n\bar{x}_n^2 - 2n\bar{x}_n x_n) \tag{14}$$

$$= \frac{n}{n-1}(\bar{x}_n - x_n)^2 \tag{15}$$

after which we have the recurrence formula for the sum of squares:

$$SS_n = SS_{n-1} + \frac{n}{n-1}(\bar{x}_n - x_n)^2 \tag{16}$$

Accordingly, the recurrence formula for the sample standard deviation (square root of variance) is:

$$s_1 := 0 \tag{17}$$

$$s_n = \sqrt{\frac{1}{n-1}\left((n-2)s_{n-1}^2 + \frac{n}{n-1}(\bar{x}_n - x_n)^2\right)} \tag{18}$$

In conclusion, a typical pseudocode for running calculation of $\bar{x}$ and $s$ indicators is as follows:

**Require:** `read_value`(): returns a new observation of $x$ at each call

1: $\bar{x} \leftarrow$ `read_value`() ▷ Initializations
2: $s \leftarrow 0$
3: $n \leftarrow 1$
4: **repeat** ▷ Main loop
5: $n \leftarrow n + 1$
6: $x \leftarrow$ `read_value`()
7: $\bar{x} \leftarrow \frac{1}{n}\left((n-1)\bar{x} + x\right)$ ▷ Update sample mean
8: $s \leftarrow \sqrt{\frac{1}{n-1}\left((n-2)s^2 + \frac{n}{n-1}(\bar{x} - x)^2\right)}$ ▷ Update sample std. dev.
9: Perform operations on $\bar{x}$ and $s$
10: **until** exit condition is true