Supplement

Anna Androvitsanea
anna.androvitsanea@gmail.com

April 3, 2021

Exercise 1

I want to calculate the second degree polynomial that gives a x_{n+1} value, which once added to a x array of lenght n its variance remains the same:

Given that Var and $\overline{x}$ is the variance and mean of the x array and Var' and $\overline{x'}$ is the variance and mean of the x array once x_{n+1} is appended I want that:

$$Var = Var' \qquad [1.1]$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Longrightarrow \sum_{i=1}^{n} x_i = n\overline{x} \qquad [1.2]$$

$$\begin{aligned} \overline{x'} = \frac{1}{n+1}\sum_{i=1}^{n+1} x_i = \frac{1}{n+1}\sum_{i=1}^{n} x_i + \frac{1}{n+1}x_{n+1} \overset{Eq.1.2}{\Longrightarrow} \\ \overline{x'} = \frac{n}{n+1}\overline{x} + \frac{1}{n+1}x_{n+1} \end{aligned} \qquad [1.3]$$

$$Var' = \frac{1}{n}\sum_{i=1}^{n+1}(x_i - \overline{x'})^2 = \frac{1}{n}\sum_{i=1}^{n+1}(x_i - \overline{x'})^2 \overset{Eq.1.3}{=} \frac{1}{n}\sum_{i=1}^{n+1}(x_i - (\frac{n}{n+1}\overline{x} + \frac{1}{n+1}x_{n+1})^2 =$$

$$\frac{1}{n}\sum_{i=1}^{n+1}(x_i - \frac{n}{n+1}\overline{x} - \frac{1}{n+1}x_{n+1})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{n}{n+1}\overline{x} - \frac{1}{n+1}x_{n+1})^2 + \frac{1}{n}(x_{n+1} - \frac{n}{n+1}\overline{x} - \frac{1}{n+1}x_{n+1})^2 =$$

$$\frac{1}{n}\sum_{i=1}^{n}[x_i^2 + (\frac{n}{n+1}\overline{x})^2 + (\frac{1}{n+1}x_{n+1})^2 - 2x_i\frac{n}{n+1}\overline{x} + 2\frac{n}{n+1}\overline{x}\frac{1}{n+1}x_{n+1} - 2x_i\frac{1}{n+1}x_{n+1}] +$$

$$\frac{1}{n}(\frac{n}{n+1}x_{n+1} - \frac{n}{n+1}\overline{x})^2 =$$

$$\frac{1}{n}\sum_{i=1}^{n}[\frac{1}{(n+1)^2}x_{n+1}^2 + (\frac{2n\overline{x}}{(n+1)^2} - \frac{2x_i}{n+1})x_{n+1} + x_i^2 + (\frac{n}{n+1}\overline{x})^2 - 2x_i\frac{n}{n+1}\overline{x}]$$

$$+\frac{n}{(n+1)^2}(x_{n+1}^2 - 2x_{n+1}\overline{x} + \overline{x}^2) =$$

$$[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(n+1)^2}]x_{n+1}^2 + \frac{1}{n}\sum_{i=1}^{n}[\frac{2n\overline{x}}{(n+1)^2} - \frac{2x_i}{n+1}]x_{n+1} + \frac{1}{n}\sum_{i=1}^{n}[x_i^2 + (\frac{n}{n+1}\overline{x})^2 - 2x_i\frac{n}{n+1}\overline{x}]$$

$$+\frac{n}{(n+1)^2}x_{n+1}^2 - \frac{2n}{(n+1)^2}\overline{x}x_{n+1} + \frac{n}{(n+1)^2}\overline{x}^2) =$$

$$[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(n+1)^2}]x_{n+1}^2 + \frac{1}{n}\sum_{i=1}^{n}[\frac{2n\overline{x}}{(n+1)^2} - \frac{2x_i}{n+1}]x_{n+1} + \frac{1}{n}\sum_{i=1}^{n}[x_i^2 + (\frac{n}{n+1}\overline{x})^2 - 2x_i\frac{n}{n+1}\overline{x}]$$

$$+\frac{n}{(n+1)^2}x_{n+1}^2 - \frac{2n}{(n+1)^2}\overline{x}x_{n+1} + \frac{n}{(n+1)^2}\overline{x}^2) =$$

$$\overset{Eq.1.1}{=} Var \Longrightarrow$$

$$[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(n+1)^2}]x_{n+1}^2 + \frac{1}{n}\sum_{i=1}^{n}[\frac{2n\overline{x}}{(n+1)^2} - \frac{2x_i}{n+1}]x_{n+1} + \frac{1}{n}\sum_{i=1}^{n}[x_i^2 + (\frac{n}{n+1}\overline{x})^2 - 2x_i\frac{n}{n+1}\overline{x}]$$

$$\frac{n}{(n+1)^2}x_{n+1}^2 - \frac{2n}{(n+1)^2}\overline{x}x_{n+1} + \frac{n}{(n+1)^2}\overline{x}^2) - Var = 0 \Longrightarrow$$

$$\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(n+1)^2} + \frac{n}{(n+1)^2}\right]x_{n+1}^2 + \frac{1}{n}\sum_{i=1}^{n}[\frac{2n\overline{x}}{(n+1)^2} - \frac{2x_i}{n+1} - \frac{2n}{(n+1)^2}\overline{x}]x_{n+1} +$$

$$\left[\frac{1}{n}\sum_{i=1}^{n}[x_i^2 + (\frac{n}{n+1}\overline{x})^2 - 2x_i\frac{n}{n+1}\overline{x}] + \frac{n}{(n+1)^2}\overline{x}^2) - Var\right] = 0 \quad [1.4]$$

I managed to bring the equation in the form $\alpha x_{n+1}^2 + \beta x_{n+1} + c = 0$ where $a = a_1 + a_2$, $b = b_1 + b_2 + b_3$ and $c = c_1 + c_2 + c_3 + c_4$.

- $a_1 = [\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(n+1)^2}]$
- $a_2 = \frac{n}{(n+1)^2}$
- $b_1 = \frac{2n\overline{x}}{(n+1)^2} - \frac{2x_i}{n+1}$
- $b_2 = \frac{1}{n}[\sum_{i=1}^{n} b_1]x_{n+1} = \frac{1}{n}\sum_{i=1}^{n}[\frac{2n\overline{x}}{(n+1)^2} - \frac{2x_i}{n+1}]$
- $b_3 = -\frac{2n}{(n+1)^2}\overline{x}$
- $c_1 = x_i^2 + (\frac{n}{n+1}\overline{x})^2 - 2x_i\frac{n}{n+1}\overline{x}$

- $c_2 = \frac{1}{n}\sum_{i=1}^{n} c_1 = \frac{1}{n}\sum_{i=1}^{n}[x_i^2 + (\frac{n}{n+1}\overline{x})^2 - 2x_i\frac{n}{n+1}\overline{x}]$
- $c_3 = \frac{n}{(n+1)^2}\overline{x}^2$
- $c_4 = -Var$

For this case I will calculate the roots and take the positive value of the result. This number will be inserted to the dataset and, as proven at Eq. 1.4calculated, will not change the variance of the column's data.