A Constant-State Method for Calculating Mean and Standard Deviation

Austin Parker [*]
austinjp@cs.umd.edu

December 21, 2005

1 Intro

Here I lay out a simple method for calculating standard deviation. Unlike the traditional method of using the set of all samples to compute the standard deviation, a method which requires one to save all seen samples before computing standard deviation, this method which has constant state. One need not save the actual values of the previous samples to compute the new average and standard deviation. This is intended for situations where one has a stream of samples and neither the state nor the computation time to store those samples.

It should be noted that the time-complexity of this method is no better than simply storing the samples and then computing the average and standard deviation from the data. The only win is in the fact that we no longer need to store all samples in order to compute the mean standard deviation.

2 Problem, Solution, Proof

Problem description: we have μ, σ, the average and standard deviation for n samples. Given s, another sample from the same population, compute μ' and σ' for those $n+1$ samples.

The algorithm is to first compute μ':

$$\mu' = \frac{n \cdot \mu + s}{n+1}$$

And then compute σ':

$$\sigma' = \sqrt{\frac{n \cdot \sigma^2 + n \cdot \left(\frac{s-\mu}{n+1}\right)^2 + (s-\mu')^2}{n+1}}$$

We now prove the above equations to give the correct answer.

---

[*]With thanks to Rob Sherwood.

Proof. That μ' is correct is trivial.

Proving σ' is trickier. By definition, we want σ' s.t.

$$\sigma'^2 = \frac{\sum_{i=1}^n (x_i - \mu')^2 + (s - \mu')^2}{n+1}$$

where x_i is the i^{th} sample. With the following algebra, we show there to be a closed form for σ' computable without access to each x_i. For reference, we use

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}, \qquad \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\begin{aligned}
\sigma'^2 &= \frac{\sum_{i=1}^n (x_i - \mu')^2 + (s - \mu')^2}{n+1} &(1)\\
&= \frac{\sum_{i=1}^n (x_i^2 - 2 \cdot x_i \cdot \mu' - \mu'^2) + (s - \mu')^2}{n+1} &(2)\\
&= \frac{\sum_{i=1}^n \left(x_i^2 - 2 \cdot x_i \cdot \frac{n \cdot \mu + s}{n+1} + \left(\frac{n \cdot \mu + s}{n+1}\right)^2\right) + (s - \mu')^2}{n+1} &(3)\\
&= \frac{\sum_{i=1}^n \left(x_i^2 - 2 \cdot x_i \cdot \frac{(n+1) \cdot \mu + s - \mu}{n+1} + \left(\frac{(n+1) \cdot \mu + s - \mu}{n+1}\right)^2\right) + (s - \mu')^2}{n+1} &(4)\\
&= \frac{\sum_{i=1}^n \left(x_i^2 - 2 \cdot x_i \cdot \mu - 2 \cdot x_i \frac{s-\mu}{n+1} + \left(\mu + \frac{s-\mu}{n+1}\right)^2\right) + (s - \mu')^2}{n+1} &(5)\\
&= \frac{\sum_{i=1}^n \left((x_i^2 - 2 \cdot x_i \cdot \mu + \mu^2) - 2 \cdot x_i \frac{s-\mu}{n+1} + 2\mu\frac{s-\mu}{n+1} + \left(\frac{s-\mu}{n+1}\right)^2\right) + (s - \mu')^2}{n+1} &(6)\\
&= \frac{\sum_{i=1}^n \left((x_i - \mu)^2 - 2 \cdot x_i \frac{s-\mu}{n+1} + 2\mu\frac{s-\mu}{n+1} + \left(\frac{s-\mu}{n+1}\right)^2\right) + (s - \mu')^2}{n+1} &(7)\\
&= \frac{\left(\sum_{i=1}^n (x_i - \mu)^2\right) - 2 \cdot n \cdot \frac{s-\mu}{n+1} \sum_{i=1}^n \left(\frac{x_i}{n}\right) + 2 \cdot n \cdot \mu\frac{s-\mu}{n+1} + n \cdot \left(\frac{s-\mu}{n+1}\right)^2 + (s - \mu')^2}{n+1} &(8)\\
&= \frac{n \cdot \sigma^2 - 2 \cdot n \cdot \frac{s-\mu}{n+1}\mu + 2 \cdot n \cdot \mu\frac{s-\mu}{n+1} + n \cdot \left(\frac{s-\mu}{n+1}\right)^2 + (s - \mu')^2}{n+1} &(9)\\
&= \frac{n \cdot \sigma^2 + n \cdot \left(\frac{s-\mu}{n+1}\right)^2 + (s - \mu')^2}{n+1} &(10)
\end{aligned}$$

□