Section 2 Report:

Streaming Computations

Question 4

Data files are provided and we had to compute certain information about the data.

Part A:

We wrote a program to compute the mean using constant amount of memory. To compute the mean(μ), we used the following formula:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

this mean value is written in a file `mean.txt` as it is useful for computations involved in the subsequent parts of the question.

Part B:

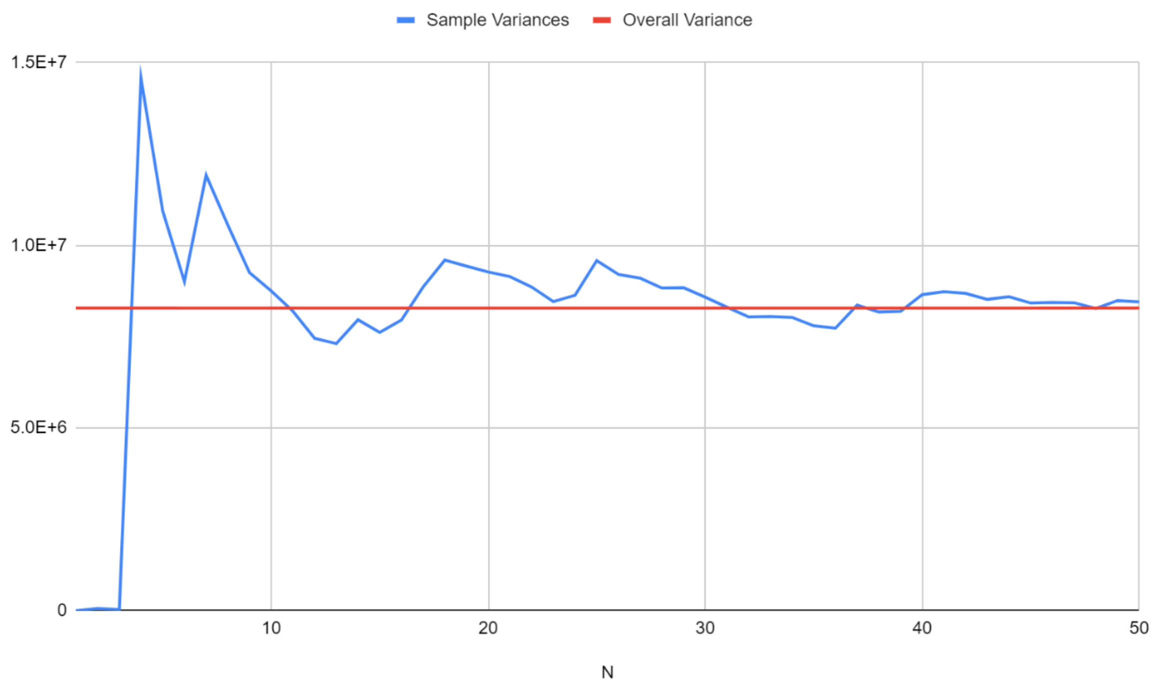
Then, we wrote a program to compute the variation of the entire data using the mean obtained and stored in `mean.txt` in Part A. As this is population variance, the following formula was used for calculation:

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

In that program, we also calculate the actual variance for each value from 1 to N to check what trend the relation between the two variances was following. To compute the running variances, the following variation of the formula for sample variance was used:

$$\sigma^2{}_i = \frac{\sum x^2 - \frac{(\sum x)^2}{i}}{i-1}$$

The running variances were then plotted against the complete population variance and the graph is observed.



So, it is clearly visible that as our code saw more and more data, the actual running variances started getting much closer to the overall approximate variance.

All the values we obtained from our code are available [here](#).

Part C:

We wrote a program to compute the percentage of numbers (x_i) which fell between 0.8μ and 1.2μ(including both 0.8μ and 1.2μ). The following formula was used.

$$Required\ percentage = \left(\frac{Number\ of\ x_i\ \in\ [0.8\mu, 1.2\mu]}{N}\right) \times 100\ \%$$
