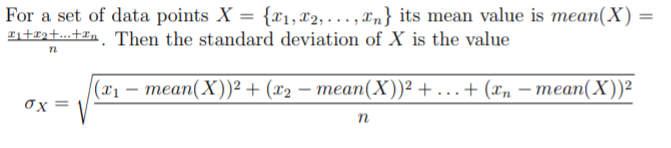
**Report**

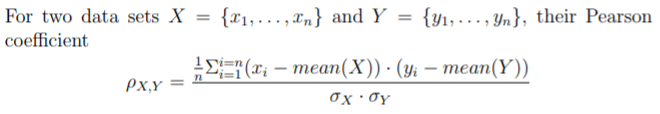
**The problem:**

The input data are given in two arrays a and b of values of type double and the length of each array is at least 2,000,000.

1.Use serial implementation without parallelism that implements the algorithm for computing Pearson correlation coefficient of data presented in arrays a and b;

2.Use MPI implementation for computing the same, that works on a varying number of MPI processes





**Where the parallelism exists:**

When calculate the summations of x[i]-mean(X), y[i]-mean(Y) and (x[i]-mean(X))\*(y[i]-mean(Y)) , I had used the parallelism MPI\_Reduce to get the answers.

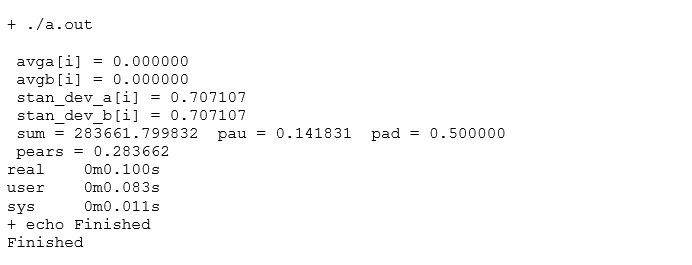
**Brief explanation of serial & MPI codes**

Serial codes: I defined three function “mean, calculateSD, pearson” and transferred them in the main function to calculate the average values, sums and Pearson coefficient of two arrays. In the main function, I defined the two arrays and populated with sin(i) and sin(i+5).

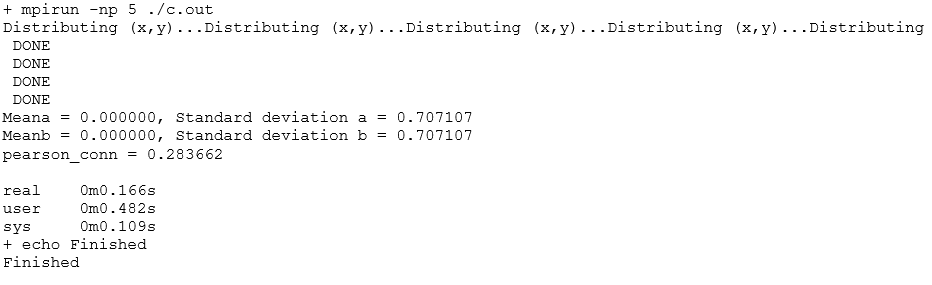
MPI codes: Firstly, I defined the two arrays and populated them with sin(i) and sin(i+5) on the root process, then initialed the MPI and used the communicator MPI\_Scatter to distribute the arrays to other processes. Also, I used the MPI\_Reduce to sum the arrays and collected the results on the root process. Finally, if the process is the root process, output the results.

**Quantified the accuracy, speed and parallel efficiency of the MPI code**

**When I use the serial code to solve the problem, the computer uses about 0.1s and the results are:**

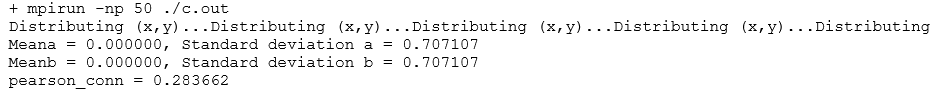


**When I use the MPI code (use 5 codes) to calculate the results, the results are:**

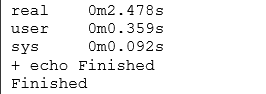


**So, Speedup S = T(serial) / T(parallel) ≈0.6, Efficiency E = T(serial) / (p\*T(parallel)) ≈0.12;**

**When I use 50 cores to compute the same problems, the results are:**



And the time are:



**So, Speedup S = T(serial) / T(parallel) ≈2.017, Efficiency E = T(serial) / (p\*T(parallel)) ≈0.043;**

**Observation:**

According to the results, I found that the T(parallel), S and E depend on cores p, and S increases with the improve of numbers of processes or cores, while the E decreased. So, using the parallelism system to solve the problems may improve the performance of computer.