**GROUP NUMBER:** 007

**GROUP MEMBERS:** Alberto Pasqualetto, Michela Schibuola, Michele Sprocatti

**TEST 1:** The goal is to test the performance of **Reservoir Sampling** (**RS**) for different values of the frequency threshold and, consequently, different sample sizes. The experiments must be run with fixed n = 1000000 and with varying values of , as indicated in the table below, using the stream emitted from port 8886. You must run each test three times, and report the values of each run as well as the average values. For each run you must report the number of estimated frequent items returned by RS and the number of true frequent items among the estimated frequent items. You must fill in the following table.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Test Reservoir Sampling, port=8886, n=1000000** | | | | | | | | | |
|  | **Number of estimated frequent items returned by RS**  **Run 1** | **Number of true frequent items returned by RS**  **Run 1** | **Number of estimated frequent items returned by RS**  **Run 2** | **Number of true frequent items returned by RS**  **Run 2** | **Number of estimated frequent items returned by RS**  **Run 3** | **Number of true frequent items returned by RS**  **Run 3** | **Average number of estimated frequent items returned by RS** | | **Average number of true frequent items returned by RS** |
| 0,07 | 10 | 8 | 11 | 6 | 11 | 8 | | 10.66 | 7.33 |
| 0,06 | 11 | 7 | 13 | 7 | 10 | 8 | | 11.33 | 7.33 |
| 0,05 | 15 | 9 | 12 | 8 | 13 | 9 | | 13.33 | 8.66 |
| 0,04 | 14 | 7 | 13 | 9 | 13 | 9 | | 13.33 | 8.33 |

**TEST 2:** The goal is to test the performance of **Sticky Samplin**g (**SS**) with increasing values of . The values of n, , and are fixed (n = 1000000, , and the experiments must be done using the stream emitted from port 8888. You must run each test three times, and report the values of each run as well as the average values. For each run you must report the number of estimated frequent items returned by SS and the number of true frequent items among the estimated frequent items. You must fill in the following table.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Test Sticky Sampling, port=8888, n=1000000, phi=0.07, delta=0.1** | | | | | | | | |
|  | **Number of estimated frequent items returned by SS**  **Run 1** | **Number of true frequent items returned by SS**  **Run 1** | **Number of estimated frequent items returned by SS**  **Run 2** | **Number of true frequent items returned by SS**  **Run 2** | **Number of estimated frequent items returned by SS**  **Run 3** | **Number of true frequent items returned by SS**  **Run 3** | **Average number of estimated frequent items returned by SS** | **Average number of true frequent items returned by SS** |
| 0,06 | 14 | 10 | 13 | 10 | 14 | 10 | 13.66 | 10 |
| 0,05 | 13 | 10 | 14 | 10 | 14 | 10 | 13.66 | 10 |
| 0,04 | 12 | 10 | 14 | 10 | 13 | 10 | 13 | 10 |
| 0,03 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |