**GROUP NUMBER: 014**

**GROUP MEMBERS: Umberto Salviati – David Petrovic – Federico Luisetto**

**TEST 1:** The goal is to assess the accuracy of the count-sketch estimations as the number of distinct items (regulated by the interval [left,right]) varies. The values D, W, and K are fixed. You must fill in the following table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ACCURACY WITH RESPECT TO NUMBER OF DISTINCT ITEMS, USING D=9, W=30, K=10**  Use 4 decimal digits for floating points | | | | |
| **[left,right]** | **Number of received distinct items in [left,right]** | **Average relative error for items with top-K frequencies** | **True normalized F2** | **Approximate normalized F2** |
| [1,15000] | 15000 | 0.0309 | 0.3749 | 0.3752 |
| [1,10000] | 10000 | 0.0085 | 0.5302 | 0.5302 |
| [1,5000] | 5000 | 0.0100 | 0.6222 | 0.6221 |
| [1,1000] | 1000 | 0.0217 | 0.4965 | 0.4966 |

**TEST 2:** The goal is to assess the accuracy of the count-sketch estimations as the number W of columns of the sketch varies. The values D, K and the interval are fixed. Repeat each experiment 3 times. You must fill in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| **ACCURACY WITH RESPECT TO NUMBER OF COLUMNS W, USING D=9, K=30, [left,right]=[1,10000]**  Use 4 decimal digits for floating points and report averages over 3 runs | | | |
| **W** | **Average relative error for items with top-K frequencies. RUN 1** | **Average relative error for items with top-K frequencies. RUN 2** | **Average relative error for items with top-K frequencies. RUN 3** |
| 100 | 0.0703 | 0.0601 | 0.0453 |
| 50 | 0.0495 | 0.0688 | 0.0821 |
| 20 | 0.8767 | 0.3830 | 1.3539 |
| 15 | 2.8344 | 1.8689 | 1.6513 |

**TEST 3:** The goal is to assess the accuracy of the count-sketch estimations as K varies. The values D, W and the interval are fixed. Repeat each experiment 3 times. You must fill in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| **ACCURACY WITH RESPECT TO K, USING D=9, W=100, [left,right]=[1,10000]**  Use 4 decimal digits for floating points and report averages over 3 runs | | | |
| **K** | **Average relative error for items with top-K frequencies. RUN 1** | **Average relative error for items with top-K frequencies. RUN 2** | **Average relative error for items with top-K frequencies. RUN 3** |
| 10 | 0.0056 | 0.0082 | 0.0043 |
| 50 | 0.0714 | 0.1473 | 0.2172 |
| 100 | 0.3213 | 0.2550 | 1.1076 |
| 200 | 1.2318 | 2.1167 | 1.3076 |