

Quiz 12 Streaming

1) Explain how reservoir sampling works with a proof by induction and an example? (3pts)

2) Explain how to calculate the false positive rate for the Bloom filter and how to find the optimal number of hash function given we known the length of the array and the number of objects in the set (3 pts)

3) Explain why the Flajolet-Martin algorithm works by showing 2^R will not be too small or too large comparing to the number of unique values in the data stream (3 pts)

4) Explain why we want to compute moments for data stream (1 pt)