

# On Range Summary Queries

Tuesday, July 11, 2023 3:30 PM (20 minutes)

Peyman Afshani, Pingan Cheng, Aniket Basu Roy and Zhewei Wei

**Abstract:** We study the query version of the heavy hitter and approximate quantile problems.

In the former problem, the input is a parameter  $\varepsilon$ , and a set  $P$  of  $n$  points in  $R^d$  where each point is assigned a color from a set  $C$  and the goal is to

build a structure such that given any geometric range  $\gamma$ ,

we can efficiently find a list of heavy hitters in  $\gamma \cap P$ ,

i.e., colors that appear at least  $\varepsilon|\gamma \cap P|$  times in  $\gamma \cap P$

as well as their approximate frequencies with an additive error of  $\varepsilon|\gamma \cap P|$ .

In the latter problem, each point is assigned a weight from a totally ordered universe

and the query must output a sequence  $S$  of  $1/\varepsilon$  weights

such that the  $i$ -th weight in  $S$  has approximate rank  $i\varepsilon|\gamma \cap P|$ , meaning,

rank  $i\varepsilon|\gamma \cap P|$  up to an additive error of  $\varepsilon|\gamma \cap P|$ .

Previously, optimal results were only known for the 1D version of the problem [WY11] but

a few sub-optimal methods were available in higher dimensions [AW17, ACH+12].

We study the problems for two important classes of geometric ranges: 3D halfspace and 3D dominance queries.

It is known that many other important queries can be reduced to these two, namely,

1D interval stabbing or interval containment,

2D three-sided queries, 2D circular as well as 2D  $k$ -nearest neighbors queries.

We consider the real RAM model of computation where integer registers of size  $w$  bits,

$w = \Theta(\log n)$ , are also available.

For dominance queries, we show optimal solutions for both heavy hitter and approximate

quantile problems: using linear space, we can answer both queries in time  $O(\log n + 1/\varepsilon)$ .

Note that as the output size is  $\frac{1}{\varepsilon}$ , after investing the initial  $O(\log n)$  searching time,

our structure takes on average  $O(1)$  time to find a heavy hitter or a quantile!

For more general halfspace heavy hitter queries, the same optimal query time can be achieved by increasing the

space by an extra  $\log_w \frac{1}{\varepsilon}$  (resp.  $\log \log_w \frac{1}{\varepsilon}$ ) factor in 3D (resp. 2D).

By spending extra  $\log^{O(1)} \frac{1}{\varepsilon}$  factors in both time and space,

we can also support quantile queries.

We remark that it is hopeless to achieve a similar query bound for dimensions 4

or higher unless significant advances are made in the data structure side of

theory of geometric approximations.

**Presenter:** CHENG, Pingan

**Session Classification:** Track A-3