

# Summaries for 2-Dimensional Orthogonal Approximate Range Counting

February 18, 2012

**Problem definition** Consider a point set  $D$  of size  $n$  on a  $u \times u$  grid. We want to build a data structure (summary) such that given a orthogonal rectangle range  $R$ , we can estimate the number of points in  $R \cap D$  with error  $\varepsilon n$ . We are interested in the size of the data structure.

**$\varepsilon$ -approximation** An  $\varepsilon$ -approximation for the point set  $D$  is a subset  $P$  such that for any rectangle range  $R$ , we have

$$\left| \frac{|R \cap P|}{|P|} - \frac{|R \cap D|}{|D|} \right| \leq \varepsilon.$$

It follows that

$$\left| \frac{n \cdot |R \cap P|}{|P|} - |R \cap D| \right| \leq \varepsilon n.$$

So given an  $\varepsilon$ -approximation  $P$  for  $D$ , we can use  $\frac{n \cdot |R \cap P|}{|P|}$  to estimate the count  $|R \cap D|$ .

The best constructive  $\varepsilon$ -approximation for 2-dimensional orthogonal rectangles has size  $O(\frac{1}{\varepsilon} \log^4 \frac{1}{\varepsilon})$  [5]. Matoušek proved a better nonconstructive bound of  $O(\frac{1}{\varepsilon} \log^{2.5} \frac{1}{\varepsilon})$  [3]. The best known lower bound is  $\Omega(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$  [2].

**Construction based on  $\varepsilon$ -net** An  $\varepsilon$ -net for the point set  $D$  is a subset  $P$ , such that for any orthogonal rectangle  $R$  with  $|R \cap D| \geq \varepsilon n$ , there is at least one point in  $R \cap P$ . For 2-dimensional orthogonal rectangles, there exists an  $\varepsilon$ -net of size  $O(\frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon})$  [1]. A matching lower bound is proved by Pach and Tardos [4].

Based on the  $\varepsilon$ -net, we have a two upper bounds. The first construction uses an  $\varepsilon$ -net of size  $O(\frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon})$ , and each point in the  $\varepsilon$ -net is associated with a structure of  $O(\frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon})$  bits. The total number of bits used is

$$O\left(\frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon} \left(\log u + \frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon}\right)\right).$$

The second construction uses an  $\frac{\varepsilon}{\log \frac{1}{\varepsilon}}$ -net of size  $O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon})$ , and each point in the  $\frac{\varepsilon}{\log \frac{1}{\varepsilon}}$ -net is associated with a structure of  $O(\log \frac{1}{\varepsilon})$  bits. The total

number of bits used is

$$O\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon} \left(\log u + \log \frac{1}{\varepsilon}\right)\right).$$

**A trivial lower bound of  $\Omega(\frac{1}{\varepsilon} \log u)$  bits** We group the  $n$  points into  $1/\varepsilon$  “fat points”, each consisting of  $\varepsilon n$  points. We place the fat points on the  $u \times u$  grid. The number of different configurations is  $u^{2/\varepsilon}$ , since each fat point can be placed in  $u^2$  positions. For two different configurations  $C_1$  and  $C_2$ , there exists a position  $(x, y)$ , such that the numbers of fat points in this position are different in  $C_1$  and  $C_2$ . This implies that given a rectangle containing only  $(x, y)$ , the range counts returned by the summaries for  $C_1$  and  $C_2$  must be different. So the summaries for the two configurations must be different, and the information-theoretic lower bound for representing the  $u^{2/\varepsilon}$  configurations is  $\Omega(\frac{1}{\varepsilon} \log u)$  bits.

**Conjecture** The first upper bound tells us that we cannot get a much higher lower bound than  $\Omega(\frac{1}{\varepsilon} \log u)$  bits if we do not impose any constraint on  $u$ . So we set  $u = n$ , and  $\varepsilon = \log n/n$ . Under this setting, we conjecture that if a summary can answer 2-dimensional orthogonal range counting query with abs error at most  $\log n$ , then it must use  $\Omega(n \log n)$  bits. In other words, answering 2-dimensional orthogonal range counting query with error  $\log n$  is as hard as answering the query with no error. It is also interesting to close the gap between the upper bound and lower bound of  $\varepsilon$ -approximation for 2-dimensional orthogonal rectangles.

## References

- [1] B. Aronov, E. Ezra, and M. Sharir. Small-size  $\varepsilon$ -nets for axis-parallel rectangles and boxes. *SIAM Journal on Computing*, 39(7):3248–3282, 2010.
- [2] J. Matoušek. *Geometric Discrepancy*. Springer, Heidelberg, Germany, 1999.
- [3] J. Matoušek. On the discrepancy for boxes and polytopes. *Monatshefte für Mathematik*, 127(4):325–336, 1999.
- [4] J. Pach and G. Tardos. Tight lower bounds for the size of epsilon-nets. In *Proceedings of the 27th annual ACM symposium on Computational geometry*, pages 458–463. ACM, 2011.
- [5] J. Phillips. Algorithms for  $\varepsilon$ -approximations of terrains. *Automata, Languages and Programming*, pages 447–458, 2008.