Good query performance with high probability

Up to now:

Fixed input set $S_n$, $n \approx (S_n)$, $\text{expected MAP}$

$n$ possible query structures $D_{n}(S_n)$ for any parameter $n$

Any $|D_{n}(S_n)|$ between $O(n)$ and $O(n^2)$

Query cost between $O(\log n)$ and $O(n)$

- Array size is $O(n)$
- For fixed query point $q$, array query cost $O(\log n)$

- Array height of $D_{n}(S_n)$ (in any structure a key point?)
- How many $D_{n}(S_n)$ form reasonable query? (for any query point)
- Worst case query time? (Construction of reasonable $D_{n}(S_n)$?)

Use: Markov Inequality

$Z \geq 0$ random variable, $\alpha > 0$

$$P(Z \geq \alpha) \leq \frac{E(Z)}{\alpha}$$

Modelling with independent random variables
Construct graph that represent all invalid orders for $S_1, S_2, S_3, S_4, \ldots, S_4$.

For any fixed invalid order, there is exactly one path $\pi$. For example, $3, 2, 4, 1$. How many paths?
fixed query point \( q \)

main edge if insertion of \( 5 \) days

the current trapezoid tool contains \( q \)

Max 4 edge segments define the current trapezoid \( T \)

Backward analysis: At most 4 segments
day the trapezoid if they are removed from a subset.

\( \Rightarrow \) Any vertex has \( \leq 4 \) marked edges from below

Now: If less than 4, simply mark some other arbitrarily.

(First three layers, note all!) Forget the case of the marks.

Analyze expected number of steps during which the
trapezoid containing \( q \) changes.

Expected number of marked edges on a source to sink path
is 6.

Random variable (choose one of the \( n \) paths)

\[ X_i = \begin{cases} 1 & \text{i-th edge is marked (root)} \\ 0 & \text{otherwise} \end{cases} \]

\( \text{No. independent mark at level } \) does not influence
mark at level \( g \) (random ports, top down? probability to hit one of the 4 segments)
2. \[ P \left[ x_i = 1 \right] = \sum_{i=1}^{4} \frac{1}{i} \quad i \geq 3 \]  

Probability that node of the form \( v \) is not hit by one of the four seques

3. \[ Y := \sum_{i=1}^{n} x_i \]  

\( Y \) is the expected number of nodes in the second part of the path to \( q \)  

\( Y \) is equal to the product of the height of the tree \( h \) and \( n \)  

\[ \sum_{i=1}^{n} x_i = n \times h \]  

\( h \) is the height of the tree

Probability that path has more than \( 2n \) height

\[ P \left[ Y \geq 2 \ln(n+1) \right] = P \left[ e^{Y} \geq (n+1)^{2t} \right] \]  

\( t > 0, \forall t \)  

\[ P \left[ \text{length searchtree to } q \geq 3 \times \ln(n+1) \right] \leq \frac{E(e^{Y})}{(n+1)^{2t}} \]  

\( E(e^{Y}) = n+1 \)  

\( t = \frac{\ln \frac{3}{4}}{\ln(n+1)} \)
\[ E(e^{tY}) = E\left(p + \sum_{i=1}^{n} x_i \right) \]
\[ = E\left( \prod_{i=1}^{n} e^{t x_i} \right) \]
\[ = \prod_{i=1}^{n} E\left(e^{t x_i}\right) \quad \text{due to independence} \]
\[ = \prod_{i=1}^{n} e^{t \cdot \frac{y_i}{\lambda} + \epsilon \left(\frac{1}{\lambda} - \frac{y_i}{\lambda}\right)} \]
\[ = \prod_{i=1}^{n} \frac{e^{t \cdot \frac{y_i}{\lambda}}}{\lambda} \]
\[ \Rightarrow \]
\[ E(x_i) = \lambda \frac{y_i}{\lambda} + 0 \cdot \left(\lambda - \frac{y_i}{\lambda} \right) \]
\[ = \prod_{i=1}^{n} \frac{y_i}{\lambda} \frac{1}{\lambda} + \frac{1-y_i}{\lambda} \]
\[ = \prod_{i=1}^{n} \frac{2}{\lambda} \frac{3}{2} \cdots \frac{n+1}{n} \]
\[ = n! \lambda \]
Lemma 50: Let $S$ be a set of $n$ non-crossing line segments, let $q$ be a query point and $k$ be a parameter $k > 0$. The probability that the second path for $q$ in $D(s)$ (random incremental construction) has more than $32 \ln(n+1)$ steps is at most $\frac{1}{(n+1)^2 \ln \frac{3}{4} - 1}$.

Proof: Illustrated.

Stated for single second path?

Maximum length of a second path $k$?

Height of $D_k(s)$?

Problem too many query points?

Collect points with the same second path in $D_k$?

Sub/strip method: vehicle lives through return
intersected by line segments
at most $2(n+1)^2$ cells
For any cell the second path is the same
for all points in the cell.
\[ P \left[ \text{height}(D_n(S_n)) \geq 32 \ln(n+1) \right] \]
\[ \leq P \left[ \text{one of the } (2n+1)^2 \text{ search paths has height } \geq 32 \ln(n+1) \right] \]
\[ \leq 2(n+1)^2 P \left[ \text{fixed search path has height } \geq 32 \ln(n+1) \right] \]
\[ \leq \frac{2(n+1)^2}{(n+1)2^{\ln \frac{5}{4}} - 1} = 2 \frac{1}{(n+1)2 \ln \frac{5}{4} - 3} \]

**Lemma 50**

**Lemma 51** Same preconditions as in Lemma 50.

The probability that the maximum couple of a search path in \( D(S_n) \) (random walk) has a height greater than 32 \( \ln(n) \) is at most

\[ \frac{2}{(n+1)2 \ln \frac{5}{4} - 3} \]

**Proof:** Just given.

**Consequence:** Choose \( z = 20 \)

\[ \ln \frac{5}{4} \approx 0.223 \]

\[ \Rightarrow P \left[ \text{height} \left( D(S_n) \right) \geq 60 \ln(n+1) \right] \leq \frac{2}{(n+1)1.4} < \frac{1}{4} \]

For \( n > 4 \).
\[
\Rightarrow \quad P\left[ \text{height } (D_{\mathcal{S}_n}) \leq 60 \ln(1 + t) \right] \geq \frac{3}{4}
\]

Similar arguments:

\[
\Rightarrow P\left[ \text{size } (D_{\mathcal{S}_n}) \leq 2n \right] \geq \frac{3}{4}
\]

\[
\Rightarrow P\left[ \text{construction cost } (D_{\mathcal{S}_n}) \leq 2n \log n \right] \geq \frac{3}{4} \quad \Rightarrow \quad \frac{3}{4} > \frac{1}{2}
\]

(max legit ad search path)

(max size)

Similar arguments also:

\[
\Rightarrow P\left[ \text{cost, cost } (D_{\mathcal{S}_n}) \leq 2n \log n \right] \geq \frac{3}{4}
\]

\[
\text{Theorem 5.2: } S \text{ plane subdivision of } n \text{ edges.}
\]

There exists a point location data structure for \( S \) of size \( \bigOmega(n) \) with \( \bigO(n \log n) \) query time. Can be computed in \( \bigO(n \log n) \) expected time.

\[
\text{Proof: Build }
\]

\[
\text{Choose random permutation of } \mathcal{S}_n \\
\text{Build: } D_{\mathcal{S}_n} \\
\text{If choice is bad, start again!}
\]

In the average 4 attempts are enough.