CS6100 – Topics in Design and Analysis of Algorithms

Beyond Worst Case Analysis

Instructor:

Raghavendra Rao B V (BSB 354, bvrr@cse.iitm.ac.in, rkunjathaya@gmail.com)

Lecture 3

Average Case Analysis

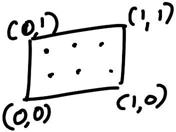
A $t(\cdot)$ F_n : inputs of length n.

$$\text{avg}_A(n) = \frac{1}{|F_n|} \sum_{x \in F_n} t(x)$$

F_n is infinite
eg $\mathbb{R}$

$$\mathbb{E}_{x \sim F_n}[t(x)]$$

k-means, ETSP : input : n points from $[0,1]^2$ $\mathbb{R}^2$



distribution: Sample $p_1, \cdots, p_n$ from $[0,1]^2$ uniformly at random.

t_{2opt} $\qquad$ $\mathbb{E}_{p_1 \cdots p_n \in [0,1]^2} \, t_{2opt}(p_1, \cdots p_n)$

t_{kmeans} $\qquad$ $\mathbb{E} \, t_{kmeans}(p_1, \cdots p_n)$

Discrete setting

e.g: Graph algorithms

$G_{n,p}$: Erdös Renyi model

$\binom{n}{2}$

$\forall\, i,j \quad i<j$

include edge (i,j) with prob p.

t_A

avg_A : $\mathbb{E}_{G \sim G_{n,p}} [t_A(G)]$

Average Case analysis: ETSP

$TSP : ([0,1]^2)^* \to \mathbb{R}$.

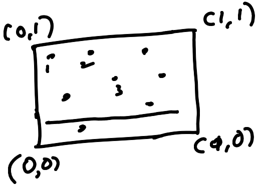
$\downarrow$ $TSP = (TSP_n)$

$TSP_n : ([0,1]^2)^n \to \mathbb{R}$

$p_1 \dots p_n$

p_i — p_j

$p_1 \dots p_n$ - uniformly independently at random from $[0,1]^2$.



$[0,1]$

Thm $\exists$ a constant c s.t. for n "sufficiently large"

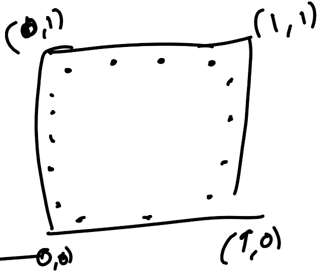
$$\lim_{n\to\infty} \mathop{E}_{p_1 \dots p_n \in [0,1]^2} \left[TSP_n(p_1, \dots p_n) \right] = c \cdot \sqrt{n}$$

$\hookrightarrow$ in fact $TSP_n(p_1 \dots p_n) = c\sqrt{n}$ with prob $1 - o(1)$

$o(1) = 0$ as $n \to \infty$

$1 - \frac{1}{2n}$

Average Case analysis: ETSP



$\underline{\underline{O(1)}}$

$\sqrt{n}$

$\underline{\underline{\text{Graph setting}}}$

$p = \frac{1}{2}$ $\quad p = \frac{1}{3}$ $\quad p = \frac{1}{100}$, $\quad p = \frac{1}{\sqrt{n}}$

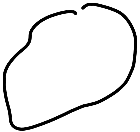
$\underline{\underline{G_{n,p}}}$

Connectivity

$\exists\, c = 1 + \epsilon$

s.t. $p < \frac{c}{n}$

then G is not con.



a

Random Graphs

Limitations of Average Case Analysis

$\exists$ $c = 1+\epsilon$ s.t

if $p < \frac{c}{n}$ then $G \sim G_{n,p}$ is not CONN with prob $1-o(1)$

& $p > \frac{c}{n}$ then G is CONN with prob $1-o(1)$

$p = \frac{c}{n}$

Threshold phenomenon

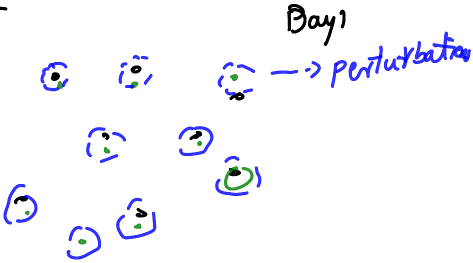
Modeling Real World ~~performance~~ inputs

Spielman - Teng (2002)

$\forall x$

$N(x)$

$\Downarrow$ neighborhood of x



Smoothed complexity at x

$$= \mathop{\mathbb{E}}_{y \in N(x)} [f(y)]$$

Perturbation Model : continuous setting

Perturbation Model : Discrete setting

Smoothed Complexity

A single step Model

Tools

- It is all about analyzing the algorithm on a distribution of inputs!!
- Typically the analysis involves identifying properties of inputs that make the algorithm run faster (or output better quality)
- Then show that for any input x, a perturbation of x satisfies this property with a good probability (or even high probability)