5

Tuesday, October 8, 2019 2:05 PM

## Average-Case Analysis

Inputs come from a probabalistic distribution D.

D models realistic inputs

Problem: Cannot predict user input well

Then: Expected / average time complexity : $\mathbb{E}_{x \sim D}[t_A(x)]$

## Randomized Algorithms

The algorithm A makes random choices r : running time is random

Worst case expected running time : $\max_{x \text{ for size } n} \mathbb{E}_r[t_A(x,r)]$ } x: inputs, r: random choices

## Universal Hashing

Choose a uniformly random hash function h from a "universal family" H.

H is a set of functions from U to $\{0, \ldots, m-1\}$

Then, use that h for all Insert / Delete / Find ops.

Heuristic : If $S \subseteq U$ are the keys in the table, $|S| = n$
choose m such that $\alpha = \frac{m}{n} = \Theta(1)$.

(A) Def: A family of hash functions H is universal if :
$\forall x, y \in U, \; x \neq y$ such that:

$$\frac{|\{h \in H : h(x) = h(y)\}|}{|H|} \leq \frac{1}{m}$$

$$\underset{h \sim H}{P}\left( h(x) = h(y) \right)$$ where $h$ uniformly picked from $h$.

(B) Thm: Suppose $H$ is universal, and $h \sim H$ uniformly at random.
then, $S = \{x_1, \dots, x_n\}$ is inserted into $T$ using $h$.

Then: $\forall x \in S$, $\mathbb{E}[n_{h(x)}] \leq 1 + \frac{n}{m}$

$\forall x \notin S$: $\mathbb{E}[n_{h(x)}] \leq \frac{n}{m}$

where $n_a$ is the number of elements in bin $a$.

Suppose that any $h \in H$ can be evaluated in $O(1)$.
then, for any sequence of $n$ inserts and $f$ finds / deletes:

$$\mathbb{E}[\text{total time}] = O\left(n + f\left(1 + \frac{n}{m}\right)\right) = \underline{O(n + f(1+\alpha))}$$

so if $\alpha = O(1)$, then expectation is $O(1)$.

In other words: for:

Find$(x)$ $\quad O(1 + n_{h(x)})$

since $n_{h(x)} \leq 1 + \alpha$ so $O(1+\alpha)$

then: $$\mathbb{E}\left[\sum_{i=1}^{f} t_i\right] = \sum_{i=1}^{f} \mathbb{E}(t_i) = \sum_{i=1}^{f} O(1+\alpha) = O(f \cdot (1+\alpha))$$

---

Justification for (b)

For any $x, y \in U$, define indicator $I_{x,y} = \begin{cases} 1 & \text{if } h(x) = h(y) \\ 0 & \text{otherwise} \end{cases}$

Define $Z_x = \left| \{ y \in S : y \neq x,\ h(y) = h(x) \} \right|$

we know $Z_x = \begin{cases} n_{h(x)} - 1 & \text{if } x \in S \\ n_{h(x)} & x \notin S \end{cases}$

Number of element collisions for things inserted

$$Z_x = \sum_{\substack{y \in S \\ x \neq y}} I_{xy} \Rightarrow \mathbb{E}[Z_x] = \mathbb{E}\Big[\sum_{\substack{y \in S \\ x \neq y}} I_{xy}\Big]$$

$$= \sum_{\substack{y \in S \\ x \neq y}} \mathbb{E}[I_{x,y}]$$

$$\mathbb{E}[I_{x,y}] = 1 \cdot P(I_{x,y} = 1) + 0 \cdot P(I_{x,y} = 0)$$

$$= P(I_{x,y} = 1)$$

$$= \sum_{\substack{y \in S \\ x \neq y}} P(h(x) = h(y)) \boxed{\leq \frac{n}{m}}$$

By definition of universal hash family summed over $n$ times

---

Constructing $H$

$U = \{0, \ldots, N-1\}$ and take prime $p \geq N$ Sample between $N$ and $2N$

$\mathbb{Z}_p = \{0, \ldots, p-1\}$ and $\mathbb{Z}_p^* = \{1, \ldots, p-1\}$

Define hash function, for any $a \in \mathbb{Z}_p^*$ and $b \in \mathbb{Z}_p$:

$$h_{a,b}(x) = \big((ax+b) \bmod p\big) \bmod m$$

Thm: $H$ is universal.

Define $g_{a,b}(x) = (ax+b) \bmod p$

Lemma. Suppose $x \neq y$, $x, y \in \mathbb{Z}_p$. The mapping from

$\{(a,b),\ a \in \mathbb{Z}_p^*,\ b \in \mathbb{Z}_p\}$ to $\{(r,s),\ r,s \in \mathbb{Z}_p,\ r \neq s\}$

given by $r = g_{a,b}(x)$, $s = g_{a,b}(y)$ is bijective.

$\Rightarrow$ If we take $a \in \mathbb{Z}_p^*$ and $b \in \mathbb{Z}_p$ uniformly at random,

$\Rightarrow$ If we take $a \in \mathbb{Z}_p^*$ and $b \in \mathbb{Z}_p$ uniformly at random, then $(g_{a,b}(x), g_{a,b}(y))$ is uniform in $\{r, s \in \mathbb{Z}_p, r \neq s\}$

$h_{a,b}(x) = g_{a,b}(x) \bmod m$. then:

$$P(h_{a,b}(x) = h_{a,b}(y)) = P(r \bmod m = s \bmod m)$$

$$= \sum_{i=0}^{m-1} P(i = r \bmod m = s \bmod m)$$

$p_i = |\{r \in \mathbb{Z}_p, r \bmod m = i\}|$

$p_i \leq \lceil p/m \rceil$, $\sum p_i = p$

$$= \sum_{i=0}^{m-1} \frac{p_i}{p} \cdot \frac{p_i - 1}{p-1}$$

$$\leq \sum_{i=0}^{m-1} \frac{p_i}{p} \cdot \frac{\lceil p/m \rceil - 1}{p-1}$$

$$= \frac{\lceil p/m \rceil - 1}{p(p-1)} \sum_{i=0}^{m-1} p_i$$

$$= \frac{\lceil p/m \rceil - 1}{p-1} \leq \frac{(p+m-1)m - 1}{p-1} = \frac{1}{m}.$$