7. Hashing construction and randomized quicksort

2018年10月25日 13:54

Construction

$\mathcal{U} = \mathbb{Z}_p := \{0, \ldots, p-1\}, p$ prime. $p \geq m$

$\mathcal{H} = \left\{ h_{ab} := \big((ax+b) \ mod \ p\big) \ mod \ m \mid a \in \mathbb{Z}_p^+, b \in \mathbb{Z}_p \right\}$

WTS $\mathcal{H}$ is universal

proof Let $x \neq y \in \mathbb{Z}_p$

lemma The mapping from $\{(a,b) \mid a \in \mathbb{Z}_p^+, b \in \mathbb{Z}_p\}$ to $\{(r,s) \in \mathbb{Z}_p^2,\}$

given by $r = (ax+b) \ mod \ p, s = (ay+b) \ mod \ p$ is bijective

prove one-to-one: suppose not one-to-one, $\Rightarrow r = s$

$0 = r - s \equiv \big(a(x-y)\big) \ mod \ p$

p is prime$\Rightarrow p \mid a \ OR \ p \mid (x-y)$

while $a < p, 0 < |x-y| < p \Rightarrow$ contradiction

prove onto: $r - s \equiv a(x-y) \ mod \ p \Rightarrow a \equiv \frac{r-s}{x-y} \ mod \ p$ is the unique solution, $b \equiv (r - ax) mod \ p$

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

$= \sum_{i=0}^{m-1} \frac{p_i(p_i-1)}{p(p-1)}$ where $p_i = \big|\{r \in \mathbb{Z}_p \mid r \ mod \ m = i\}\big|, then \ \sum p_i = P, p_i \leq ceil\left(\frac{p}{m}\right)$

$$\leq \frac{ceil\left(\frac{p}{m}\right) - 1}{P(P-1)} \sum p_i \leq \frac{\frac{p+m-1}{m} - 1}{p-1} = \frac{1}{m}$$

Randomized quicksort

r_quicksort(A)

pick p uniformly random from {1, … , n}

A_< = array of all A[i] < A[p]

A_> = array of all A[i] > A[p]

r_quicksort(A_<)

r_quicksort(A_>)

A = [A<, A[p], A>]

Observations

At each call each element is compared with A[p]

Two elements are only compared at most once if one of that is the pivot

claim the WC expected runtime is in $O(n \log n)$

proof let $A[r_1, \ldots, r_n]$ be the sorted array

Define $X_{ij} := I\big(A[r_i], A[r_j] \ are \ ever \ compared\big)$

$$E(\#comparisons) = E\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(X_{ij}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(X_{ij} = 1)$$

in any call where $A[r_i], A[r_j]$ are compared together, $i < j$

consider p is picked:

1) $A[p] < A[r_i] \ AND \ A[p] > A[r_j]$, they gonna stay in the same array, either A_< or A_>
2) $A[p] = A[r_i] \ OR \ A[p] = A[r_j]$, they are compared and will never be compared
3) $A[r_i] < A[p] < A[r_j]$, they are splited into A_< and A_> and can never be compared.

For case 1, there is always a moment they will go into case 2) and 3) in the following stack call

ij

given by case 2), $P\left(X_{ij} = 1\right) = \frac{2}{j-1+1}$

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P\left(X_{ij} = 1\right) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j - i + 1} = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} \leq 2 \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{1}{k} \Rightarrow \in \Theta(n \log n)$$