Lecture 6-b
Randomized Quicksort

View in slide-show mode
Randomized Quicksort

- In the avg-case analysis, we assumed that all permutations of the input array are equally likely.
  - But, this assumption does not always hold
  - e.g. What if all the input arrays are reverse sorted?
    ➔ Always worst-case behavior

- Ideally, the avg-case runtime should be independent of the input permutation.
- Randomness should be within the algorithm, not based on the distribution of the inputs.
  i.e. The avg case should hold for all possible inputs
Randomized Algorithms

- Alternative to assuming a uniform distribution:
  ➔ Impose a uniform distribution
  e.g. Choose a random pivot rather than the first element

- Typically useful when:
  - there are many ways that an algorithm can proceed
  - but, it’s difficult to determine a way that is always guaranteed to be good.
  - If there are many good alternatives; simply choose one randomly.
Randomized Algorithms

- Ideally:
  - Runtime should be independent of the specific inputs
  - No specific input should cause worst-case behavior
  - Worst-case should be determined only by output of a random number generator.
Randomized Quicksort

Using Hoare’s partitioning algorithm:

\[
\text{R-QUICKSORT}(A, p, r)
\]

\[
\text{if } p < r \text{ then}
\]

\[
q \leftarrow \text{R-PARTITION}(A, p, r)
\]

\[
\text{R-QUICKSORT}(A, p, q)
\]

\[
\text{R-QUICKSORT}(A, q+1, r)
\]

\[
\text{R-PARTITION}(A, p, r)
\]

\[
s \leftarrow \text{RANDOM}(p, r)
\]

\[
\text{exchange } A[p] \leftrightarrow A[s]
\]

\[
\text{return } \text{H-PARTITION}(A, p, r)
\]

Alternatively, permuting the whole array would also work

\[\Rightarrow\] but, would be more difficult to analyze
Randomized Quicksort

Using Lomuto’s partitioning algorithm:

**R-QUICKSORT**\((A, p, r)\)

```plaintext
if \( p < r \) then
    q ← R-PARTITION\((A, p, r)\)
    R-QUICKSORT\((A, p, q-1)\)
    R-QUICKSORT\((A, q+1, r)\)
```

**R-PARTITION**\((A, p, r)\)

```plaintext
s ← RANDOM\((p, r)\)
return L-PARTITION\((A, p, r)\)
```

Alternatively, permuting the whole array would also work

➡ but, would be more difficult to analyze
Notations for Formal Analysis

- Assume all elements in $A[p..r]$ are distinct
- Let $n = r - p + 1$

- Let $\text{rank}(x) = |\{A[i]: p \leq i \leq r \text{ and } A[i] \leq x\}|
  
i.e. rank(x) is the number of array elements with value less than or equal to x

<table>
<thead>
<tr>
<th>p</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

$\text{rank}(5) = 3$

i.e. it is the 3rd smallest element in the array
The following analysis will be for Quicksort using Hoare’s partitioning algorithm.

Reminder: The pivot is selected randomly and exchanged with A[p] before calling H-PARTITION

Let x be the random pivot chosen.

What is the probability that rank(x) = i for i = 1, 2, …n?

\[ P(\text{rank}(x) = i) = \frac{1}{n} \]
Various Outcomes of H-PARTITION

Assume that \( \text{rank}(x) = 1 \)

\[ \text{i.e. the random pivot chosen is the smallest element} \]

What will be the size of the left partition (\(|L|\))?

\[ \Rightarrow |L| = 1 \]

\[ \text{pivot} = x = 2 \]

Reminder: Only the elements less than or equal to \( x \) will be in the left partition.
Various Outcomes of H-PARTITION

Assume that $\text{rank}(x) > 1$

i.e. the random pivot chosen is not the smallest element

What will be the size of the left partition ($|L|$)?

Reminder: Only the elements less than or equal to $x$ will be in the left partition.

Reminder: The pivot will stay in the right region after H-PARTITION if $\text{rank}(x) > 1$

$\Rightarrow |L| = \text{rank}(x) - 1$

pivot = $x = 5$
Various Outcomes of H-PARTITION - Summary

\[ P(\text{rank}(x) = i) = \frac{1}{n} \quad \text{for } 1 \leq i \leq n \]

\[ \text{if rank}(x) = 1 \text{ then } |L| = 1 \]

\[ \text{if rank}(x) > 1 \text{ then } |L| = \text{rank}(x) - 1 \]

\[ P(|L| = 1) = P(\text{rank}(x) = 1) + P(\text{rank}(x) = 2) \]

\[ P(|L| = i) = P(\text{rank}(x) = i+1) \]

\[ \text{for } 1 < i < n \]

\[ P(|L| = 1) = \frac{2}{n} \]

\[ P(|L| = i) = \frac{1}{n} \]

\[ \text{for } 1 < i < n \]

x: pivot

|L|: size of left region
Various Outcomes of H-PARTITION - Summary

<table>
<thead>
<tr>
<th>rank(x)</th>
<th>probability</th>
<th>T(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/n</td>
<td>T(1) + T(n-1) + ( \Theta(n) )</td>
</tr>
<tr>
<td>2</td>
<td>1/n</td>
<td>T(1) + T(n-1) + ( \Theta(n) )</td>
</tr>
<tr>
<td>3</td>
<td>1/n</td>
<td>T(2) + T(n-2) + ( \Theta(n) )</td>
</tr>
<tr>
<td>( i+1 )</td>
<td>1/n</td>
<td>T(i) + T(n-i) + ( \Theta(n) )</td>
</tr>
<tr>
<td>n</td>
<td>1/n</td>
<td>T(n-1) + T(1) + ( \Theta(n) )</td>
</tr>
</tbody>
</table>
Average - Case Analysis: Recurrence

\[ T(n) = \frac{1}{n} (T(1) + T(n-1)) \]
\[ + \frac{1}{n} (T(1) + T(n-1)) \]
\[ + \frac{1}{n} (T(2) + T(n-2)) \]
\[ + \frac{1}{n} (T(i) + T(n-i)) \]
\[ + \frac{1}{n} (T(n-1) + T(1)) \]
\[ + \Theta(n) \]

\( x = \text{pivot} \)

\( rank(x) \)
\[ 1 \]
\[ 2 \]
\[ 3 \]
\[ \vdots \]
\[ i+1 \]
\[ \vdots \]
\[ n \]
Recurrence

\[ T(n) = \frac{1}{n} \sum_{q=1}^{n-1} (T(q) + T(n-q)) + \frac{1}{n} (T(1) + T(n-1)) + \Theta(n) \]

Note: \[ \frac{1}{n} (T(1) + T(n-1)) = \frac{1}{n} (\Theta(1) + O(n^2)) = O(n) \]

\[ \Rightarrow T(n) = \frac{1}{n} \sum_{q=1}^{n-1} (T(q) + T(n-q)) + \Theta(n) \]

- for \( k = 1, 2, \ldots, n-1 \) each term \( T(k) \) appears twice
  once for \( q = k \) and once for \( q = n-k \)

- \[ T(n) = \frac{2}{n} \sum_{k=1}^{n-1} T(k) + \Theta(n) \]
Solving Recurrence: Substitution

**Guess:** \( T(n) = \Theta(n \lg n) \)

**I.H.:** \( T(k) \leq ak \lg k \) for \( k < n \), for some constant \( a > 0 \)

\[
T(n) = \frac{2}{n} \sum_{k=1}^{n-1} T(k) + \Theta(n)
\]

\[
\leq \frac{2}{n} \sum_{k=1}^{n-1} (ak \lg k) + \Theta(n)
\]

\[
= \frac{2a}{n} \sum_{k=1}^{n-1} (k \lg k) + \Theta(n)
\]

Need a tight bound for \( \sum k \lg k \)
Tight bound for $\sum klk$  

- Bounding the terms

$$\sum_{k=1}^{n-1} k \lg k \leq \sum_{k=1}^{n-1} n \lg n = n(n-1) \lg n \leq n^2 \lg n$$

This bound is not strong enough because

- $T(n) \leq \frac{2a}{n} n^2 \lg n + \Theta(n)$

  $$= 2an \lg n + \Theta(n) \quad \Rightarrow \text{couldn’t prove } T(n) \leq an \lg n$$
Tight bound for $\sum k \lg k$

• Splitting summations: ignore ceilings for simplicity

\[
\sum_{k=1}^{n-1} k \lg k \leq \sum_{k=1}^{n/2-1} k \lg k + \sum_{k=n/2}^{n-1} k \lg k
\]

First summation: $\lg k < \lg(n/2) = \lg n - 1$

Second summation: $\lg k < \lg n$
Splitting: \[
\sum_{k=1}^{n-1} k \log k \leq \sum_{k=1}^{n/2-1} k \log k + \sum_{k=n/2}^{n-1} k \log k
\]

\[
\sum_{k=1}^{n-1} k \log k \leq (\log n - 1) \sum_{k=1}^{n-1} k + \log n \sum_{k=n/2}^{n-1} k
\]

\[
= \log n \sum_{k=1}^{n-1} k - \sum_{k=1}^{n/2-1} k = \frac{1}{2} n(n-1) \log n - \frac{1}{2} \frac{n}{2} \left(\frac{n}{2} - 1\right)
\]

\[
= \frac{1}{2} n^2 \log n - \frac{1}{8} n^2 - \frac{1}{2} n \left(\log n - 1/2\right)
\]

\[
\sum_{k=1}^{n-1} k \log k \leq \frac{1}{2} n^2 \log n - \frac{1}{8} n^2 \quad \text{for } \log n \geq 1/2 \Rightarrow n \geq \sqrt{2}
\]
Substituting: \[ \sum_{k=1}^{n-1} k \log k \leq \frac{1}{2} n^2 \log n - \frac{1}{8} n^2 \]

\[ T(n) \leq 2a \left( \frac{1}{n^2} \sum_{k=1}^{n} k \log k \right) + \left( \frac{1}{n^2} \right) + \left( \frac{1}{8} n^2 \right) + \left( n \right) \]

\[ = an \log n + \frac{a}{4} n \left( n \right) \]

We can choose \( a \) large enough so that \( a \frac{1}{4} n \left( n \right) \)

\[ T(n) = O(n \log n) \]

Q.E.D.