Exit Tickets

average ⇒ distribution of inputs

If random pivot can be bad, why do it?

$\boxed{\text{A)}}$ Simple + fast way to choose pivot. Usually not unlucky

B) In practice, we always pick pivot to be median
O(nlogn)

---

random aspect to alg   deterministic

When average + when worst-case?

Expectation Value vs. Probability of Unlucky?

Why use QuickSort if can be slow?

Partition → See review video

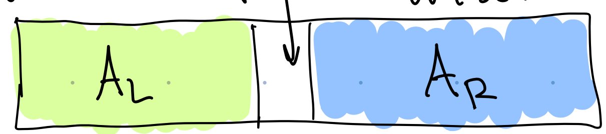
What is "Expectation Value?"

QuickSort

Input: Array A of unique integers
Output: Sorted A

- If $|A| = 1$: Return A
- pivInd $\leftarrow$ randomly chosen index, with value pivVal } Preprocessing
- Partition (A, pivInd) } Dividing
- QuickSort (A_L)
- QuickSort (A_R) } Conquering

After Partition:

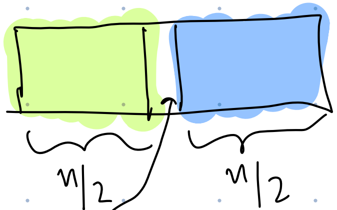


Key Pts

- Partition takes $O(|A|)$ time
- If pivVal is z_i (i^{th} smallest element of A) after Partition, pivVal is at position i.

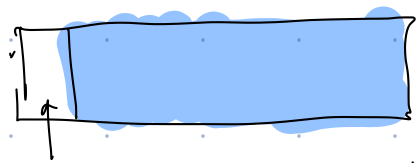
Lucky vs. Unlucky Pivot Choices

Lucky:



pivVal is median array value

Unlucky:



pivVal is smallest/largest array value

1. Suppose you get lucky at every recursive call of QuickSort.
2. Suppose you get unlucky at every recursive call of QuickSort.

- Create recurrence relation for runtime of QuickSort in each case
- Solve recurrence to determine runtime in each case

3. What is Sample Space? Random Variable? Expectation value? Linearity of Expectation?

Lucky vs. Unlucky Pivot Choices

1. Suppose you get lucky at every recursive call or QuickSort.

$$T(n) = \begin{cases} O(1) & \text{if } n \leq 1 \\ \underset{\uparrow\, A_L}{T\left(\frac{n}{2}\right)} + \underset{\uparrow\, A_R}{T\left(\frac{n}{2}\right)} + O(n) = 2T\left(\frac{n}{2}\right) + O(n) \end{cases} \overset{\text{Tree formula}}{\Big\}} \underbrace{O(n\log n)}_{\text{Great!}}$$

2. Suppose you get unlucky at every recursive call of QuickSort.

$$T(n) = \begin{cases} O(1) & n \leq 1 \\ T(n-1) + O(n) \end{cases} \Rightarrow \text{Expand + Hope} \quad \underbrace{T(n) = O(n^2)}_{\text{Awful}}$$

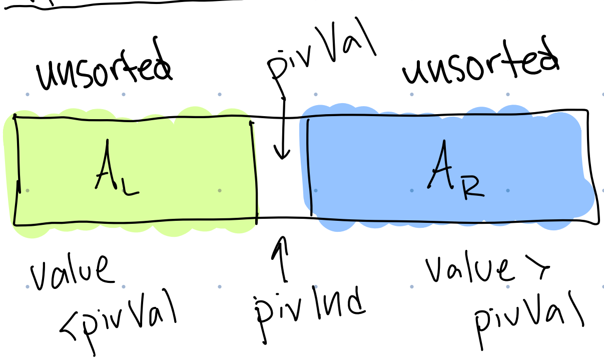
Partition(A, pivInd, pivVal)

- Swap pivot with A[1]
- Current ← 2
- While current ≤ |A|:
    - ☆ If A[current] < pivVal:
        - Swap A[current], pivVal
        - Swap A[pivInd+1], pivVal
    - current ++

← every element of array is compared to pivot

Strategy: count the # of times ☆ is run over the whole algorithm

After Partition:



Analyzing Average Runtime

1. Determine Sample Space S = set of all possible sequences of random events that might occur over the course of alg.

   ex: QuickSort → S = set of possible sequences of pivot choices that the alg might make

What is the sample space if QuickSort is run on

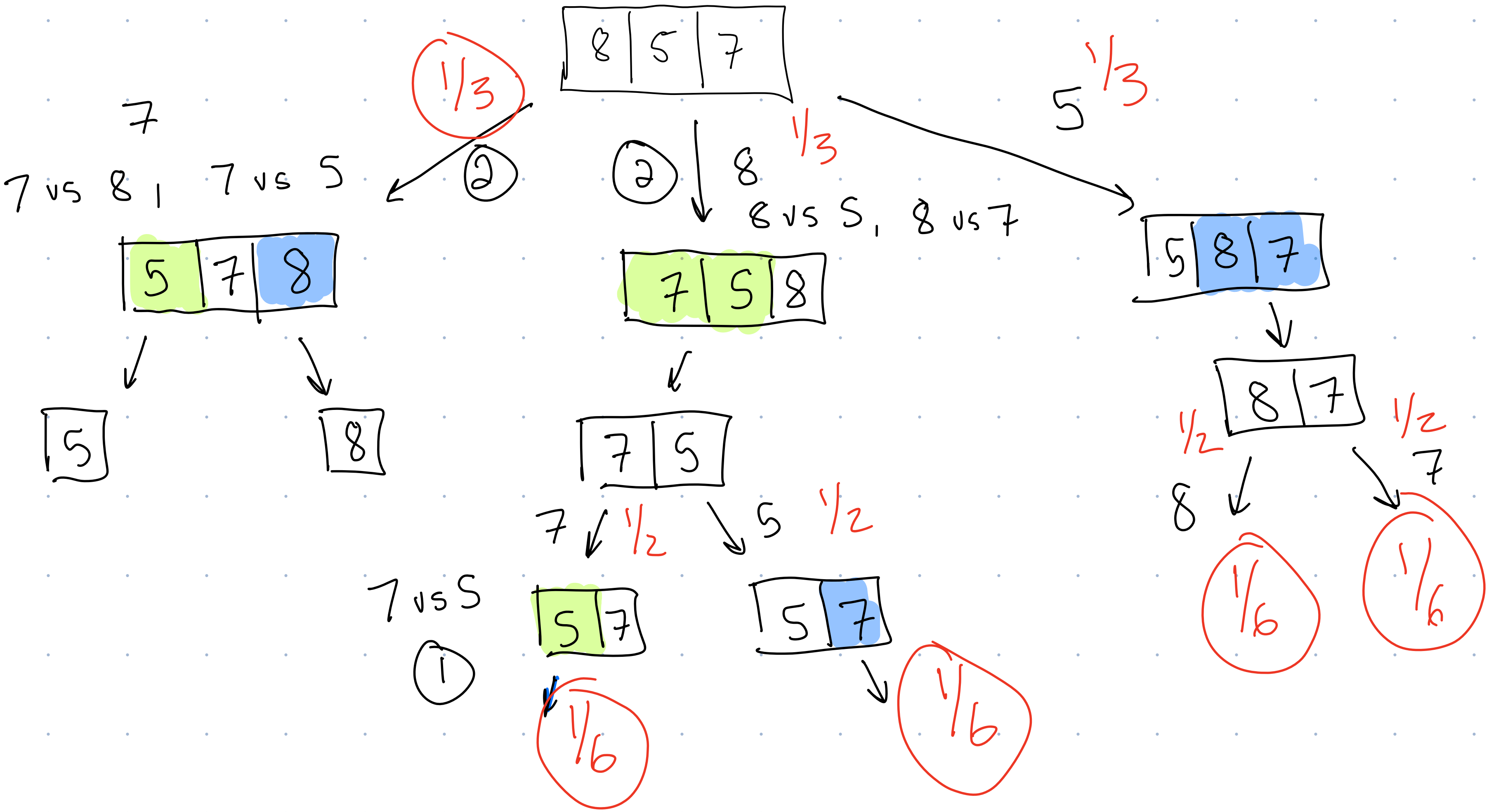
8	5	7

A) $S = \{8, 5, 7\}$

B) S = All possible permutations of $\{8, 5, 7\}$

C) S = Power set of $\{8, 5, 7\}$ (set of all subsets of $\{8, 5, 7\}$)

$\boxed{\text{D)}}$ $S = \{(7), (8, 5), (8, 7), (5, 8), (5, 7)\}$



$$S = \{7, (8,7), (8,5), (5,8), (5,7)\}$$

$$R(7) = 2 \qquad R(8,7) = 3$$

Analyzing Average Runtime

2. Create a Random Variable that maps each element of the sample space to a number

$R : S \to \mathbb{R}$

\# of times ✱ is run

ex: QuickSort: $R(\sigma) = \boxed{\#\text{ of comparisons}}$ of 2 elements of our array if pivot sequence σ is chosen.

3. Take Expectation value of R to get average run time

prob. of σ occuring

$$\boxed{\mathbb{E}[R] = \sum_{\sigma \in S} p(\sigma) \cdot R(\sigma)}$$

ex: $\boxed{8|5|7}$ (7) (5 7) (5 8) (8 5) (8 7)

$= \frac{1}{3} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 3 = 2\frac{2}{3}$

2. (Alternate) $R: S \to \mathbb{R}$ $\Rightarrow$ break up into a sum of simple random variables

$f(x) = x + x^2$ $\quad$ $f_1(x) = x$

$f = f_1 + f_2$ $\quad$ $f_2(x) = x^2$

$X_{ij}(\sigma)$ = # of comparisons between i^{th} smallest element of A (z_i) and the j^{th} smallest element of A (z_j)

8	5	7

$z_1 = 5$ $\quad$ $z_2 = 7$ $\quad$ $z_3 = 8$

$X_{23}((8,5)) = 1$ (# of times 7 and 8 are compared over the course of alg if pivot choices are 8, 5)

$$R(\sigma) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}(\sigma) \iff$$ total # of comparisons done over course of alg.

$$= \sum_{i<j} X_{ij}$$

4 $$\mathbb{E}[R] = \mathbb{E}\left[\sum_{i<j} X_{ij}\right]$$

$$= \sum_{i<j} \mathbb{E}[X_{ij}]$$ linearity of Expectation

To Analyze $E[X_{ij}]$, consider:

- Pivot in recursive call

- In Partition, pivot is compared to every other elt, and that is it.

- Suppose z_i, z_j $(i<j)$ are both in a subarray that is input to some recursive call of QuickSort. For each of the following cases (*)
  - are z_i, z_j compared in this call?
  - are they kept together or separated in future recursive calls

  * z_i or z_j chosen as pivot
  * z_k chosen as pivot
    * $k > i, j$
    * $k < i, j$
    * $i < k < j$

- What values can X_{ij} take (only 2 possible), and under which conditions does it take those values?

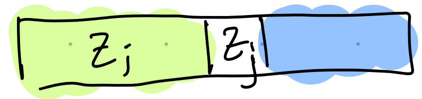
- What is probability of z_i, z_j being compared?

To Analyze $\mathbb{E}[X_{ij}]$, consider:

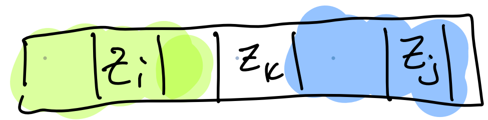
1 ☆ z_i or z_j chosen as pivot (z_i)

- z_i, z_j are compared
- Separated and will never be compared again

$X_{ij} = 1$



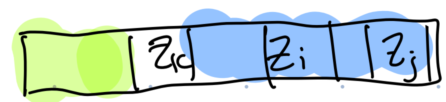
2 ☆ z_k chosen as pivot, $i < k < j$



- z_i, z_j not compared
- Separated

$X_{ij} = 0$

3 ☆ z_k chosen as pivot, $k < i, j$



- z_i, z_j not compared
- Kept together

X_{ij} not decided, might be compared or not in future

$X_{ij} \in \{0, 1\}$ so X_{ij} indicator random variable

Back to Average Runtime:

$$\mathbb{E}[R(\sigma)] = \mathbb{E}\Big[\sum_{i<j} X_{ij}\Big] = \sum_{i<j} \mathbb{E}[X_{ij}]$$

$$\mathbb{E}[X_{ij}] = \sum_{\sigma \in S} \Pr(\sigma) \cdot X_{ij}(\sigma)$$

$$= \underbrace{\sum_{\substack{\sigma \in S: \\ X_{ij}(\sigma)=0}} \Pr(\sigma)\, \cancelto{0}{X_{ij}(\sigma)}}_{0} + \sum_{\substack{\sigma \in S: \\ X_{ij}(\sigma)=1}} \Pr(\sigma) \cdot \cancelto{1}{X_{ij}(\sigma)}$$

$$= \sum_{\substack{\sigma \in S \\ X_{ij}(\sigma)=1}} \Pr(\sigma)$$

by def

$= \text{Probability that } X_{ij} = 1$

$= \text{Probability that } z_i, z_j \text{ are compared}$

Probability that $X_{ij} = 1$

comparison $X_{ij} = 1$

$z_1 \quad z_2 \quad z_3 \quad \cdots \quad z_{i-1} \quad z_i \quad z_{i+1} \quad \cdots \quad z_{j-1} \quad z_j \quad z_{j+1} \quad \cdots \quad z_n$

Pivot here: delayed decision

No comparison $X_{ij} = 0$

Pivot here: delayed decision

$j - i + 1$

$\frac{2}{j-i+1}$

What is the probability that z_i, z_j are compared?

A) $\frac{1}{j-i}$

$\boxed{\text{B) } \frac{2}{j-i+1}}$

~~C)~~ $\frac{2}{n}$

~~D)~~ $\frac{1}{n^2}$

$n = j \quad i = 1 \quad \rightarrow \quad \frac{2}{n}$

$j = i + 1 \quad \frac{2}{2}$

Continuing $\mathbb{E}[R]$ analysis:

$$\mathbb{E}[R] = \sum_{i<j} \mathbb{E}[X_{ij}] = \sum_{i<j} \Pr(z_i, z_j \text{ are compared})$$

$$= \sum_{i<j} \frac{2}{j-i+1}$$

$$= \sum_{i=1}^{n-1} \underbrace{\sum_{j=i+1}^{n} \frac{2}{j-i+1}}$$

$$\frac{2}{(i+1)-i+1} = \frac{2}{2} \qquad \frac{2}{n-i+1}$$

$$= \sum_{i=1}^{n-1} \left[\frac{2}{2} + \frac{2}{3} + \frac{2}{4} + \cdots \frac{2}{n-i+1} \right]$$

$$\leq \sum_{i=1}^{n} \left[\frac{2}{1} + \frac{2}{2} + \frac{2}{3} + \frac{2}{4} + \cdots \frac{2}{n-i+1} + \frac{2}{n-i+2} + \frac{2}{n-i+3} + \cdots \frac{2}{n} \right]$$

$$= \sum_{i=1}^{n} \left(2 \underbrace{\sum_{j=1}^{n} \frac{1}{j}}_{\text{math fact}} \right)$$

$$\leq \sum_{i=1}^{n} 2(\ln(n) + 1)$$

$$= 2n(\ln(n+1)) = O(n \log n)$$