Analysis of
Randomized Quicksort

Randomized Quicksort

- <u>Quiz 1</u>: Given an array of n ≥ 4 distinct elements $a_1 < a_2 < \ldots < a_n$, what is the probability that a_3 and a_4 are compared during randomized quicksort?

A. 0

B. 1/n

C. 2/n

D. 1

Randomized Quicksort

- Quiz 2: Given an array of n ≥ 4 distinct elements $a_1 < a_2 < \ldots < a_n$, what is the probability that a_2 and a_4 are compared during randomized quicksort?

A. 0

B. 2/n

C. 1/3

D. 2/3

Randomized Quicksort

- Quiz 3: Given an array of $n \geq 2$ distinct elements $a_1 < a_2 < \ldots < a_n$, what is the probability that a_1 and a_n are compared during randomized quicksort?

A. 0

B. 1/n

C. 2/n

D. 1

Randomized Quicksort

- Quiz 4: Given an array of $n \geq 10$ distinct elements $a_1, a_2, \ldots, a_n$, what is the probability that a_4 and a_{10} are compared during randomized quicksort?

Analysis of Randomized Quicksort

Let RV X = number of comparisons over all calls to Partition.

Suffices to compute $E[X]$.

Analysis of Randomized Quicksort

Let RV X = number of comparisons over all calls to Partition.

Suffices to compute $E[X]$.

Notation:

- Let $z_1, z_2, \ldots, z_n$ denote the list items (in sorted order).
- Let $Z_{ij} = \{z_i, z_{i+1}, \ldots, z_j\}$.

Let RV $X_{ij} = \begin{cases} 1 & \text{if } z_i \text{ is compared to } z_j \\ 0 & \text{otherwise} \end{cases}$

That is, X_{ij} is an **indicator random variable**.

$X_{ij}=I\{z_i \text{ is compared to } z_j\}$.

Thus,

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}.$$

Analysis (Continued)

We have:

$$E[X] = E\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right]$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[X_{ij}]$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P[z_i \text{ is compared to } z_j]$$

So, all we need to do is to compute $P[z_i$ is compared to $z_j]$.

Analysis (Continued)

z_i and z_j are compared iff the first element to be chosen as a pivot from Z_{ij} is either z_i or z_j. [note: $Z_{ij} = \{z_i, z_{i+1}, \ldots, z_j\}$]

Exercise: discuss this.

$$
\begin{aligned}
P[z_i \text{ is compared to } z_j] &= P[z_i \text{ or } z_j \text{ is first pivot from } Z_{ij}] \\
&= P[z_i \text{ is first pivot from } Z_{ij}] \\
&\quad + P[z_j \text{ is first pivot from } Z_{ij}] \\
&= \frac{1}{j-i+1} + \frac{1}{j-i+1} \\
&= \frac{2}{j-i+1}
\end{aligned}
$$

Analysis (Continued)

Therefore,

$$E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}$$

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

$$< \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k}$$

$$= \sum_{i=1}^{n-1} O(\lg n)$$

$$= \boxed{O(n \lg n).}$$

Substitute $k = j - i$.

$$\sum_{k=1}^{n} \frac{1}{k} = H_n \text{ (n}^{\text{th}} \text{ Harmonic number)}$$

$$H_n = \ln n + O(1), \text{ Eq. (A.7)}$$

Exercises on indicator random variables

- Text p 122: 5.2-1, 5.2-2, 5.2-3, 5.2-4, 5.2-5

Deterministic vs. Randomized Algorithms

- Deterministic Algorithm : Identical behavior for different runs for a given input. Worst case analysis.
- Randomized Algorithm : Behavior is generally different for different runs for a given input. Average runtime analysis.
  - Better: compute expectation & variance