Quicksort

\textsc{Quicksort}(A, p, r)

1 \quad \textbf{if} \; p < r
2 \quad \textbf{then} \; q \leftarrow \textsc{Partition}(A, p, r)
3 \quad \textsc{Quicksort}(A, p, q - 1)
4 \quad \textsc{Quicksort}(A, q + 1, r)

\textsc{Partition}(A, p, r)

1 \quad y \leftarrow \textsc{Random}(p, r)
2 \quad \textbf{Exchange} \; A[y] \; \textbf{and} \; A[r]
3 \quad x \leftarrow A[r]
4 \quad i \leftarrow p - 1
5 \quad \textbf{for} \; j \leftarrow p \; \textbf{to} \; r - 1
6 \quad \quad \textbf{do} \; \textbf{if} \; A[j] \leq x
7 \quad \quad \quad \textbf{then} \; i \leftarrow i + 1
8 \quad \quad \textbf{exchange} \; A[i] \leftrightarrow A[j]
9 \quad \textbf{exchange} \; A[i + 1] \leftrightarrow A[r]
10 \quad \textbf{return} \; i + 1
Partition Loop Invariant

\textsc{Partition}(A, p, r)

1. \( y \leftarrow \text{RANDOM}(p, r) \)
2. Exchange \( A[y] \) and \( A[r] \)
3. \( x \leftarrow A[r] \)
4. \( i \leftarrow p - 1 \)
5. for \( j \leftarrow p \) to \( r - 1 \)
6. \hspace{1em} do if \( A[j] \leq x \)
7. \hspace{2em} then \( i \leftarrow i + 1 \)
8. \hspace{1em} exchange \( A[i] \leftrightarrow A[j] \)
9. exchange \( A[i + 1] \leftrightarrow A[r] \)
10. return \( i + 1 \)

\textbf{Loop Invariant} \hspace{.5em} At the beginning of each iteration of the for loop in \textsc{Partition}

1. \( A[p \ldots i] \leq x \)
2. \( A[i + 1 \ldots j - 1] > x \)
3. \( A[j \ldots r - 1] \) is unexamined
4. \( A[r] = x \)
Quicksort Analysis

- $T(n)$ is the expected running time of quicksort
- Partition takes $O(n)$ time.
- If partition is $x$ th smallest, then

$$T(n) = T(x) + T(n - x) + O(n)$$
Quicksort Analysis

• $T(n)$ is the expected running time of quicksort
• Partition takes $O(n)$ time.
• If partition is the $x$th smallest, then

$$T(n) = T(x) + T(n - x) + O(n)$$

For intuition consider cases:
• $x = n/2$
• $x = n/10$
• $x = 1$
**Cases**

\[ T(n) = T(x) + T(n - x) + O(n) \]

1: \( x = n/2 \)

\[
T(n) = T(n/2) + T(n/2) + O(n) \\
= 2T(n/2) + O(n) \\
= O(n \log n)
\]

2: \( x = n/10 \)

\[
T(n) = T(n/10) + T(9n/10) + O(n) \\
= O(n \log n)
\]

3: \( x = 1 \)

\[
T(n) = T(1) + T(n - 1) + O(n) \\
= T(n - 1) + O(n) \\
= O(n^2)
\]

What might this make us guess the answer is?
Following the Selection Analysis

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

could continue as in Selection.
Alternative Analysis

• We will count comparisons of data elements.
• Claim 1: The running time is dominated by comparison of data items.
• Claim 2: All comparisons are in line 6 of Partition, and compare some item \( A[j] \) to the pivot element.
• Claim 3: Once an element is chosen as a pivot, it is never compared to any other element again.
• Claim 4: Each pair of elements is compared to each other at most once.

Analysis:
• Let the data be renamed \( Z_1, \ldots, Z_n \) in sorted order.
• Use \( Z_{ij} \) to denote \( Z_i, Z_{i+1}, \ldots, Z_j \)
• Let \( X_{ij} \) be the indicator random variable for the comparison of \( Z_i \) to \( Z_j \).
• Let \( X \) be the random variable counting the number of comparisons.
• By claim 4, we have

\[
X = \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{ij}
\]
Analysis

\[ X = \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{ij} \]

Taking expectations

\[
E[X] = E\left[ \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{ij} \right] \\
= \sum_{i=1}^{n} \sum_{j=i+1}^{n} E[X_{ij}] \quad \text{linearity of expectation} \\
= \sum_{i=1}^{n} \sum_{j=i+1}^{n} Pr(Z_i \text{ is compared to } Z_j)
\]
What is the probability that $Z_i$ is compared to $Z_j$?

When is $Z_i$ compared to $Z_j$?

- When either $Z_i$ or $Z_j$ is chosen as a pivot, and the other one is still in the same recursive problem.
What is the probability that $Z_i$ is compared to $Z_j$?

When is $Z_i$ compared to $Z_j$?

- When either $Z_i$ or $Z_j$ is chosen as a pivot, and the other one is still in the same recursive problem.
- Equivalently, when either $Z_i$ or $Z_j$ is the first element from $Z_{ij}$ to be chosen as a pivot.
- What is the probability that $Z_i$ or $Z_j$ is the first element from $Z_{ij}$ to be chosen as a pivot?
What is the probability that $Z_i$ is compared to $Z_j$?

When is $Z_i$ compared to $Z_j$?

- When either $Z_i$ or $Z_j$ is chosen as a pivot, and the other one is still in the same recursive problem.
- Equivalently, when either $Z_i$ or $Z_j$ is the first element from $Z_{ij}$ to be chosen as a pivot.

What is the probability that $Z_i$ or $Z_j$ is the first element from $Z_{ij}$ to be chosen as a pivot?

- $Z_{ij}$ has $j - i + 1$ elements.
- pivots are always chosen uniformly at random
- $Pr(Z_i$ is compared to $Z_j) = 2/(j - i + 1)$
Finishing analysis

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

\[ = \sum_{i=1}^{n} \sum_{j=i+1}^{n} E[X_{ij}] \quad \text{linearity of expectation} \]

\[ = \sum_{i=1}^{n} \sum_{j=i+1}^{n} Pr(Z_i \text{ is compared to } Z_j) \]

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

make the transformation of variables \( k = j - i + 1 \)

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

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

\[ \leq 2 \sum_{i=1}^{n} \ln(n - i + 1) \]

\[ \leq 2 \sum_{i=1}^{n} \ln(n) \]

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