Same start as for deterministic selection

\[
\text{SELECT}(A, i, n)
\]

1. if \((n = 1)\)
2. then return \(A[i]\)
3. \(p = A[\text{Random}(1, n)]\)
4. 5.
6. \(L = \{x \in A : x \leq p\}\)
   
   \(H = \{x \in A : x > p\}\)
7. if \(i \leq |L|\)
8. then \(\text{SELECT}(L, i, |L|)\)
9. else \(\text{SELECT}(H, i - |L|, |H|)\)

\[
\frac{1}{2} \text{ the time } \\
p \text{ is in the middle}
\]

First or last is bad

\[
|L| = 1 \quad |H| = n - 1
\]

\[
p_{\text{min}} \\
p_{\text{max}}
\]

Worst-case running time

\[
\infty \text{ running time.}
\]
Analysis

$T(n)$ is expected running time.

$T(n) = \sum_{x=1}^{n} \Pr(\text{partition is } x \text{ smallest})$ - Running time when partition is $x$ smallest

Using $x$ and $n-x$ as an upper bound of the sizes of the two sides:

$$T(n) \leq \sum_{x=1}^{n} \frac{1}{n} ((T(x) \text{ or } T(n-x)) + O(n))$$

$$\leq \sum_{x=1}^{n} \frac{1}{n} (T(\max\{x, n-x\}) + O(n))$$

$$\leq \left(\frac{1}{n}\right) \sum_{x=1}^{n} (T(\max\{x, n-x\})) + O(n)$$

We now rewrite the max term. Notice that as $x$ goes from 1 to n, the term $\max\{x, n-x\}$ takes on the values $n-1, n-2, n-3, \ldots, n/2, n/2, n/2 + 1, n/2 + 2, \ldots, n-1, n$. As an overestimate, we say that it takes all the values between $n/2$ and $n$ twice. Thus we substitute and obtain:

$T(n) = \sum_{x=1}^{n} \Pr(\text{partition is } x \text{ smallest})$ - Running time when partition is $x$ smallest

Using $x$ and $n-x$ as an upper bound of the sizes of the two sides:

$$T(n) \leq \sum_{x=1}^{n} \frac{1}{n} ((T(x) \text{ or } T(n-x)) + O(n))$$

$$\leq \sum_{x=1}^{n} \frac{1}{n} (T(\max\{x, n-x\}) + O(n))$$

$$\leq \left(\frac{1}{n}\right) \sum_{x=1}^{n} (T(\max\{x, n-x\})) + O(n)$$

We now rewrite the max term. Notice that as $x$ goes from 1 to n, the term $\max\{x, n-x\}$ takes on the values $n-1, n-2, n-3, \ldots, n/2, n/2, n/2 + 1, n/2 + 2, \ldots, n-1, n$. As an overestimate, we say that it takes all the values between $n/2$ and $n$ twice. Thus we substitute and obtain.
Using $x$ and $n - x$ as an upper bound of the sizes of the two sides:

$$T(n) \leq \sum_{x=1}^{n} \left( \frac{1}{n} \left( T(x) \text{ or } T(n-x) \right) + O(n) \right)$$

$$\leq \sum_{x=1}^{n} \left( \frac{1}{n} T(\max\{x, n-x\}) + O(n) \right)$$

$$\leq \left( \frac{1}{n} \right) \sum_{x=1}^{n} \left( T(\max\{x, n-x\}) + O(n) \right)$$

We now rewrite the max term. Notice that as $x$ goes from 1 to $n$, the term $\max\{x, n-x\}$ takes on the values $n-1, n-2, n-3, \ldots, n/2, n/2, n/2+1, n/2+2, \ldots, n-1, n$. As an overestimate, we say that it takes all the values between $n/2$ and $n$ twice. Thus we substitute and obtain

$$T(n) \leq \frac{2}{n} \sum_{x=0}^{n/2} T(n/2 + x) + O(n)$$

$$= \frac{2}{n} T(n) + \left( \frac{2}{n} \sum_{x=0}^{n/2-1} T(n/2 + x) \right) + O(n)$$

---

QuickSort

257149836
L
251437986

123456789
3. QUICKSORT(A, p, q - 1)
4. QUICKSORT(A, q + 1, r)

PARTITION(A, p, r)
1. y ← RANDOM(p, r)
2. Exchange A[y] and A[r]
3. x ← A[r]
4. i ← p - 1
5. for j ← p to r - 1
6. do if A[j] ≤ x
7. then i ← i + 1
10. return i + 1

If position is
Xth smallest T(n) = T(x) + T(n-x) + O(n)

Intuition
If x = \frac{n}{2} T(n) = T(\frac{n}{2}) + T(\frac{n}{2}) + O(n)
= 2T(\frac{n}{2}) + O(n)
= O(n \log n)

If x = \frac{n}{10} T(n) = T(\frac{n}{10}) + T(\frac{9n}{10}) + O(n)
= O(n \log n)
\[ T(n) = T(1) + T(n-1) + O(n) \]
\[ = T(n-1) + O(n) \]
\[ \sum_{i=1}^{n} \frac{1}{i} = O(n^2). \]

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

Solve this...
Count comparisons of array elements

All comparisons are from lines of PARTITION

\[ A(j) \leq x \quad x \text{ is the pivot} \]

\[ \vdash \text{all comparisons involve the pivot} \]

after one run of PARTITION, pivot is fixed in the right place.

\[ \vdash \text{each pair of elements is compared at most once.} \]

---

```
1. y ← RANDOM(p, r)
2. Exchange A[y] and A[r]
3. \( x ← A[r] \)
4. \( i ← p − 1 \)
5. for \( j ← p \) to \( r − 1 \)
6.  \[ \underline{\text{do if } A[j] ≤ x} \]
7.  \[ \quad \text{then } i ← i + 1 \]
10. return \( i + 1 \)
```

**Loop Invariant** At the beginning of each iteration of the for loop in Partition

1. \( A[p \ldots i] \leq x \)
2. \( A[i + 1 \ldots j - 1] > x \)
Let data be \( z_1, \ldots, z_n \) in sorted order

\[ Z_{ij} = \{ z_i, z_{i+1}, \ldots, z_{j-1}, z_j \} \]

Example result on \( Z_{15} \to Z_{28} \)

Let \( X_{ij} = 1 \) if \( z_i \) compared to \( z_j \)

\[ X_i = \text{total # comps.} \]

\[ X = \sum \sum X_{ij} \quad E(X) = \sum \sum E(X_{ij}) = \sum \sum \Pr(z_i \text{ is comp. to } z_j) \]

What is \( \Pr( z_i \text{ compared to } z_j ) \)?

When is \( z_i \text{ compared to } z_j \)

- either \( z_i \) or \( z_j \) is chosen as pivot
  - the other one (of \( z_i \) or \( z_j \)) is still in the same set.

\[ \iff \] in the set of number \( Z_{ij} \),

either \( z_i \) or \( z_j \) is the first pivot chosen,
out of \( Z_{ij} \)

\[ z_{10}, z_{11}, z_{12}, z_{13}, z_{14} \quad z_{25} \quad z_{26} \quad \ldots \]

If \( z_{12} \) chosen first out of \( z_{10}, z_{16} \) and \( z_{16} \) are never compared but if \( z_{16} \) or \( z_{16} \) chosen first then they are compared.
\[ A(\text{from } z_{ij} \text{ if } i < j \text{ is first chosen}) = \frac{2}{j-i+1} \]

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

\[ = \sum_{i=1}^{n} \frac{2}{k} \geq \ln(n-i+1) \leq 2 \sum_{i=1}^{n} \ln n = O(n \ln n). \]