- Recursion for randomized quicksort

\[ E(T_n) = n + \frac{1}{n} \sum_{i=1}^{n} (E(T_{i-1}) + E(T_{n-i})) \]

1. Index of pivot
2. Time to recursively right part.
3. Solve left part.

1. Guess the solution

\[ \text{observe: similar to mergesort recursion } T(n) = n + 2T(\frac{n}{2}) \]

\[ n \]

\[ i-1 \]

\[ n-i \]

guess \( E(T_n) \leq C \cdot n \log_2 n \) for some large constant \( C \)

2. Base case: define \( E(T_1) = E(T_0) = 0 \)

\[ E(T_2) = 1 \]

assume \( E(T_k) \leq C \cdot k \log_2 k \) for all \( k < n \)

\[ E(T_n) = n + \frac{1}{n} \sum_{i=1}^{n} (E(T_{i-1}) + E(T_{n-i})) \]

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

\[ \leq n + \frac{1}{n} \sum_{i=1}^{n} C \cdot i \log_2 i \]

Observation: \( \log_2 i \) is sometimes \( \leq \log_2 n \)

more precisely \( \log_2 i \leq \log_2 n - 1 \)

\[ \leq n + \frac{2}{n} \sum_{i=1}^{\lfloor n/2 \rfloor} C \cdot i \log_2 i \]

\[ \leq n + \frac{2}{n} \sum_{i=1}^{\lfloor n/2 \rfloor} C \cdot \log_2 n - \frac{2}{n} \sum_{i=1}^{\lfloor n/2 \rfloor} C \cdot i \log_2 n \]

\[ \leq n + \frac{2}{n} \cdot \frac{n(n-1)}{2} \log_2 n - \frac{2}{n} \cdot \frac{(n+1)n}{6} \cdot C \]

\[ \leq C \cdot n \log_2 n + (n - \frac{Cn}{4}) \]

(when \( C > 4 \))
- Quick Selection
  - Given an array of size $n$, find the $k$-th smallest number.
  - Idea: similar to quick sort
    - Pick a random pivot number in the array
    - Partition the array (same as quick sort)
    - Decide which part has the number we want
    - Recurse only on the relevant part
  - Let $T_n$ be the time for quick selection on $n$ numbers.
  - $T_n = n + \begin{cases} Ti - 1 & \text{if } i > k \quad (k \text{ is in the left part}) \\ 0 & \text{if } i = k \\ Ti - 1 & \text{if } i < k \quad (k \text{ is in the right part}) \end{cases}$
  - $E[T_n] = n + \frac{1}{n} \sum_{i=1}^{k-1} E[T_i - 1] + \frac{1}{n} \sum_{i=k+1}^{n} E[T_i - 1]$

- Birthday Paradox
  - What is the prob. that $k$ people all have different birthdays?
  - Define random variables
    - $A_k = 1$ if all $k$ people have different birthdays
    - $0$ otherwise (there are at least a pair that has the same birthday)
  - Want to compute $Pr[A_k = 1]$
  - Idea: use conditioning
    - $Pr[A_k = 1 | A_{k-1} = 1]$: if first $k-1$ people have different birthdays
    - What is the prob. that $k$-th person does not have the same birthday with the $k-1$ people?
    - $\Rightarrow Pr[A_k = 1 | A_{k-1} = 1] = \frac{365 - (k-1)}{365} = \frac{365 - k + 1}{365}$
    - Let $n = 365$
    - $Pr[A_k = 1 | A_{k-1} = 1] = 1 - \frac{k-1}{n}$
    - $Pr[A_k = 1 | A_{k-1} = 0] = 0$
\[ \Pr[A_k = 1] = \Pr[A_{k-1} = 1] \cdot \Pr[A_k | A_{k-1} = 1] \]
\[ = \Pr[A_{k-1} = 1] \cdot (1 - \frac{k-1}{n}) \]

Base case \( \Pr[A_1 = 1] = 1 \)

\[ \Pr[A_k = 1] = (1 - \frac{k-1}{n}) \cdot (1 - \frac{k-2}{n}) \cdot (1 - \frac{k-3}{n}) \cdots (1 - \frac{1}{n}) \cdot \Pr[A_1 = 1] \]

Observation: \( e^x \geq 1 + x \)

\[ \Rightarrow e^{-\frac{k}{n}} \geq (1 - \frac{k}{n}) \]

\[ \Pr[A_k = 1] \leq e^{-\frac{k}{n}} \cdot e^{-\frac{k-1}{n}} \cdot e^{-\frac{k-2}{n}} \cdots e^{-\frac{1}{n}} \]
\[ = e^{-\frac{k(k-1)}{2n}} \]
\[ = e^{-\frac{K(K-1)}{2n}} \]

When \( K \geq \sqrt{2n+1} \), \( \Pr[A_k = 1] \leq e^{-1} \)

- Coupon collector
  - \( n \) types of coupon, \( 1 \$ \) gives one coupon at random
  - How much money does it take to get \( \geq 1 \) of each type?
  - Define random variable
    - Let \( T_i \) be the amount of money spent after getting the \((i-1)\)-th distinct coupon, before getting the \(i\)-th coupon.

\[
\begin{array}{ccccccc}
1 & 2 & 4 & 2 & 5 & 4 & 1 & 6 \\
T_1 & T_2 & T_3 & T_4 & T_5 \\
\end{array}
\]

- Let \( T \) to be the amount of money for \( n \) distinct coupons

\[ T = T_1 + T_2 + T_3 + \ldots + T_n \]

- Linearity of expectation \( \mathbb{E}[T] = \sum \mathbb{E}[T_i] \)
- how to compute \( \mathbb{E}[T_i] ? \)

\[ \Pr[T_i = 1] = \frac{n - (i-1)}{n} = 1 - \frac{i-1}{n} \]
\[\Pr[T_i = 1] = \frac{n - (i-1)}{n} = 1 - \frac{i-1}{n}\]

\[\Pr[T_i = 2] = \frac{i-1}{n} \times \left(\frac{1}{n}\right)\]

Pr[first coupon is a duplicate] Pr[2nd coupon is new]

\[\Pr[T_i = 3] = \left(\frac{i-1}{n}\right)^2 \left(1 - \frac{i-1}{n}\right)\]

\[\Pr[T_i = t] = \left(\frac{i-1}{n}\right)^{t-1} \left(1 - \frac{i-1}{n}\right)\]

\[\text{Geometric distribution}\]

\[ECT_i = \sum_{t=1}^{\infty} t \cdot \Pr[T_i = t] = \frac{n}{n - (i-1)}\]

\[ECT = ECT_1 + ECT_2 + \ldots + ECT_n\]

\[= \frac{n}{n} + \frac{n}{n-1} + \ldots + \frac{n}{1}\]

\[= n \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}\right)\]

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