

- Randomized Quicksort (Analysis)
- Quick Selection
- Birthday Paradox
- Coupon Collector

- Recursion for randomized quicksort

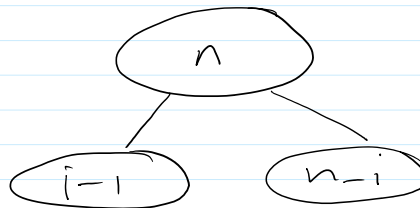
$$E[T_n] = n + \frac{1}{n} \sum_{i=1}^n (E[T_{i-1}] + E[T_{n-i}])$$

$\uparrow$ time to recursively solve left part $\uparrow$ right part.

i : index of pivot

1. Guess the solution

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



guess $E[T_n] \leq C n \log_2 n$ for some large constant C

2. induction: base case define $E[T_0] = E[T_1] = 0$

$$E[T_2] = 1$$

assume $E[T_k] \leq C k \log_2 k$ for all $k < n$

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

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

$$\leq n + \frac{2}{n} \sum_{i=1}^{n-1} C i \log_2 i$$

observation: $\log_2 i$ is sometimes $< \log_2 n$
more precisely if $i \leq \frac{n}{2}$, $\log_2 i \leq \log_2 n - 1$

$$= n + \frac{2}{n} \sum_{i=1}^{\frac{n}{2}} C i \log_2 i + \frac{2}{n} \sum_{i=\frac{n}{2}+1}^{n-1} C i \log_2 i$$

$$\leq n + \frac{2}{n} \sum_{i=1}^{\frac{n}{2}} C i (\log_2 n - 1) + \frac{2}{n} \sum_{i=\frac{n}{2}+1}^{n-1} C i \log_2 n$$

$$= n + \frac{2}{n} \sum_{i=1}^{\frac{n}{2}} C i \log_2 n - \frac{2}{n} \sum_{i=1}^{\frac{n}{2}} C i$$

$$= n + \frac{2}{n} \cdot \frac{n(n-1)}{2} C \log_2 n - \frac{2}{n} \cdot \frac{(\frac{n}{2}+1)\frac{n}{2}}{2} \cdot C$$

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

$$\leq C \cdot n \log_2 n$$

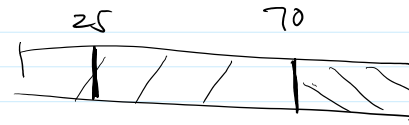
(when $C \geq 4$) $\square$

- 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)

$n=100$

$k=36$

- 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.

$$T_n = n + \begin{cases} T_{n-i} & \text{if } i < k \quad (k \text{ is in the right part}) \\ T_{i-1} & \text{if } i > k \quad (k \text{ is in the left part}) \\ 0 & i = k \end{cases}$$

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

- Birthday Paradox

- What is the prob. that k people all have different birthdays?

- Define random variables

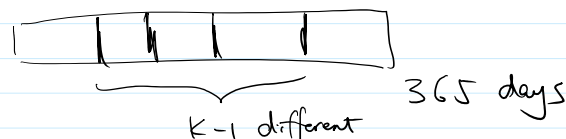
$A_k = 1$ if all k people have different birthdays

$= 0$ 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



$$\Rightarrow \Pr[A_k=1 | A_{k-1}=1] = \frac{365 - (k-1)}{365} = \left(1 - \frac{k-1}{365}\right)$$

$$\text{Let } n=365 \quad \Pr[A_k=1 | A_{k-1}=1] = 1 - \frac{k-1}{n}$$

$$\Pr[A_k=1 | A_{k-1}=0] = 0$$

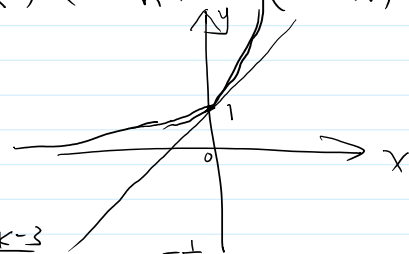
$$\begin{aligned}\Pr[A_k=1] &= \Pr[A_{k-1}=1] \cdot \Pr[A_k=1 | A_{k-1}=1] \\ &= \Pr[A_{k-1}=1] \cdot \left(1 - \frac{k-1}{n}\right)\end{aligned}$$

base case $\Pr[A_1=1] = 1$

$$\Pr[A_k=1] = \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{k-2}{n}\right) \left(1 - \frac{k-3}{n}\right) \dots \left(1 - \frac{1}{n}\right) \cdot \Pr[A_1=1]$$

observation: $e^x \geq 1+x$

$$\Rightarrow e^{-\frac{i}{n}} \geq \left(1 - \frac{i}{n}\right)$$



$$\begin{aligned}\Pr[A_k=1] &\leq e^{-\frac{k-1}{n}} \cdot e^{-\frac{k-2}{n}} \cdot e^{-\frac{k-3}{n}} \dots e^{-\frac{1}{n}} \\ &= e^{-\sum_{i=1}^{k-1} \frac{i}{n}} \\ &= e^{-\frac{k(k-1)}{2n}}\end{aligned}$$

when $k \geq \sqrt{2n} + 1$ $\Pr[A_k=1] \leq e^{-1}$

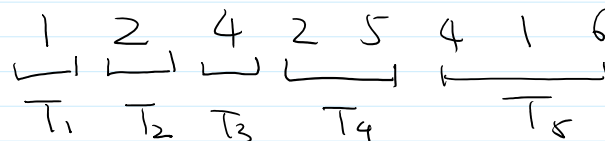
- Coupon collector

n types of coupon, 1\$ gives one coupon at random

- How much money does it take to get ≥ 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.

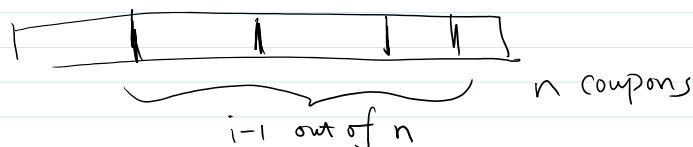


- Let T to be the amount of money for n distinct coupons

$$T = T_1 + T_2 + T_3 + \dots + T_n$$

- linearity of expectation $\mathbb{E}[T] = \sum_{i=1}^n \mathbb{E}[T_i]$

- how to compute $\mathbb{E}[T_i]$?



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

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

$$\Pr[T_i=2] = \underbrace{\frac{i-1}{n}}_{\Pr[\text{first coupon is a duplicate}]} \times \underbrace{\left(1 - \frac{i-1}{n}\right)}_{\Pr[\text{2nd coupon is new}]}$$

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

$$\Pr[T_i=t] = \left(\frac{i-1}{n}\right)^{t-1} \left(1 - \frac{i-1}{n}\right) \quad \text{geometric distribution}$$

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

$$- \mathbb{E}[T] = \mathbb{E}[T_1] + \mathbb{E}[T_2] + \dots + \mathbb{E}[T_n]$$

$$= \frac{n}{n} + \frac{n}{n-1} + \dots + \frac{n}{1}$$

$$= n \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}\right)$$

$$= \Theta(n \log n)$$

□