Randomization → Indicator variables

$$X = \begin{cases} 1 & p \\ 0 & 1-p \end{cases}$$

① $E[X] = p = 1 \cdot p + 0(1-p)$

② $E[aX + bY] = aE[X] + bE[Y]$

even if X and Y are depending on each other

RANDOM PERMUTATION

rand(a,b) chooses uniformly at random
an integer in $[a..b]$
with $pr = \frac{1}{b-a+1}$

In practice, it's pseudorandom as truly
random is undecidable (curios: Kolmogorov
complexity)

$A[1..n]$

for $i = 2, 3, \ldots, n$:
  $j = \text{rand}(i, n)$
  swap $A[i]$ and $A[j]$

use indicator variables to show that each permutation is generated with $pr = \frac{1}{n!}$

BIRTHDAY PARADOX

n days $\qquad (n = 365)$

how many people so that two of them at least have the same birthday (out of n possible ones)

Let's use indicator variables: person i, person $j \neq i$
$i < j$

$$X_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ have same birthday} \quad p \\ 0 & \text{otherwise} \quad 1-p \end{cases}$$

$$E[X_{ij}] = p = \sum_{k=1}^{n} \frac{1}{n^2} = \frac{1}{n}$$

both i and j are born on day $k \in [n]$

person i $= \frac{1}{n}$ to be born in a certain day
person j

m = # people

$$X = \sum_{i=1}^{m} \sum_{j=i+1}^{m} X_{ij}$$ = how many ordered pairs of people (among the m ones) have the same birthday

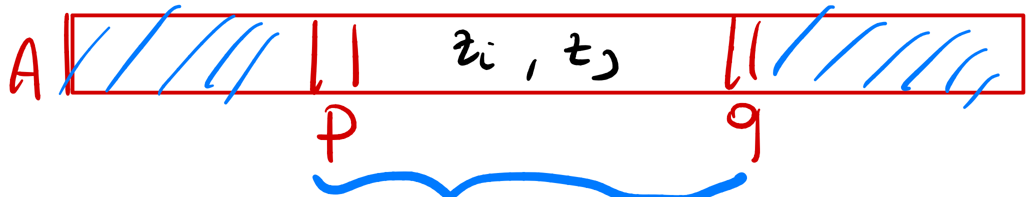
$$E[X] = \underbrace{\sum_{i=1}^{m} \sum_{j=i+1}^{m}}_{\binom{m}{2}} \underbrace{E[X_{ij}]}_{\frac{1}{n}} = \frac{m \cdot (m-1)}{2} \cdot \frac{1}{n}$$

$E[X] \geq 1$ holds when $m(m-1) \geq 2n$

$\Rightarrow m \sim \sqrt{2n}$ $\quad$ $n = 365 \Rightarrow m \sim 27$

RANDOMIZED QUICK SORT

key step in QS:



choose element to split A[p..q]

$r = \text{rand}(p, q)$

use A[r] to split PIVOT

- worst-case is $O(n^2)$ &
- average case is $O(n \lg n)$
- expected case is $O(n \lg n)$ ← much stronger notion

Analysis of RQS

1) For the purpose of analysis, output is

$$z_1 < z_2 < \cdots < z_n$$

2) When two keys $z_i : z_j$ are compared?

one of the two <u>must</u> be the current pivot

3) how many times $z_i : z_j$ are compared?

at most once

4) $z_i : z_j$ are compared :

$$|A[p..q]| \geq j - i + 1$$

because all of $z_i, z_{i+1}, \ldots, z_j$ are in $A[p..q]$

$$X_{ij} = \begin{cases} 1 & \text{if } z_i : z_j \text{ are compared at any time} \\ 0 & \text{o.w} \end{cases}$$

$i<j$

$$X = \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{ij}$$ = # comparisons because of 3)

$E[X]$?

$E[X_{ij}] = Pr(X_{ij} = 1) = Pr\left[\begin{matrix} z_i, z_j \text{ are in the same } A[p..q] \\ \underline{\text{and}} \quad z_i \text{ or } z_j \text{ is the pivot for } A[p..q] \end{matrix}\right]$

↑ x_{ij} is indicator

$\leq Pr[z_i \text{ or } z_j \text{ is the pivot for } A[p..q]] =$

$Pr[A \cap B] \leq Pr[B]$

$$\frac{2}{|A[p..q]|} \overset{4)}{\leq} \frac{2}{j-i+1}$$

$$E[X] = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \underbrace{E[X_{ij}]}_{\leq \frac{2}{j-i+1}} \leq \sum_{i=1}^{n} \underbrace{\sum_{j=i+1}^{n} \frac{2}{i+1}}_{\leq 2 \ln n} = O(n \lg n)$$