The method of Indicator Random Variable

- Simple technique for computing the expected value of a random variable:

$$E[X] = \sum_x x\,p(x)$$

- Given a sample space and an event A, an indicator random variable for event A is

$$X_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

Then $E[X_A] = 1 \cdot P(A) + 0 \cdot [1 - P(A)]$

$= P(A)$

- Together with linearity of expectation, this is powerful.

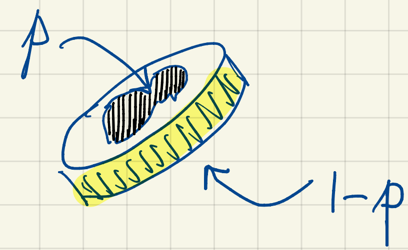
- Let's see how this is useful

<u>Example:</u> Toss the coin n times and let X be the number of heads. $X \in \{0, 1, 2, \dots, n\}$

What is $E[X]$?

$$E[X] = \sum_k k\, P(X=k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = ?$$

Let $X_i = \begin{cases} 1 & \text{Head on toss } i \\ 0 & \text{otherwise} \end{cases}$



$$E[X_i] = P(H) = p$$

Observe $X = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$

Linearity: $E[X] = E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$

$= np$ (done!)

General Strategy

1) Define indicator random variables

$$X_i = \begin{cases} 1 & A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

2) Find $E[X_i] = P(A_i)$ [should be easy]

3) Express quantity of interest X as $X = \sum_i X_i$

4) Use Linearity of expectation to find $E[X]$

$$E[X] = \sum_i E[X_i]$$

Consider hiring problem in book (slightly modified)

- We interview n candidates, this happens over time
- If the current candidate is the best so far, we hire
- There is a cost of hiring $c = 1$

```
Hire(A, n)    ▷ A is an array of distinct positive integers
    cost ← 0
    best ← 0
    for i ← 1 to n
        do if A[i] > best
            then best ← A[i]
                 ▷ hire candidate i
                 cost ← cost + 1
    return cost
```

$A[i] > A[j]$ means candidate i is more qualified than candidate j

Worst case:

When numbers appear in increasing order
i.e. candidates appear in increasing order of qualification
We hire all of them, cost is $\underline{\underline{n}}$

Probabilistic Analysis (avg. case):

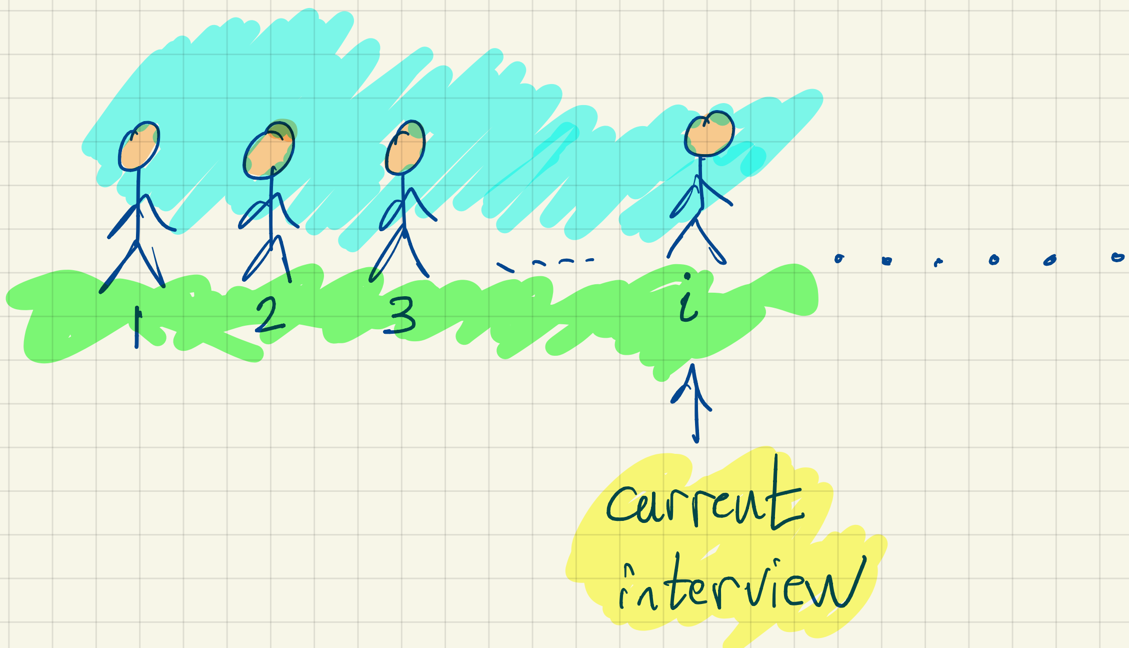
- Assume uniform random permutation

each of the $n!$ permutations appear with equal probability

- We hire candidate i if $A[i]$ is largest among $A[1 \ldots i]$
- Under above assumption, this happens with prob. $\frac{1}{i}$ since each of the first i candidates is equally likely to be the best.

(Well, this might require a proof but let's follow intuition)

Illustration



Under **uniform random permutation**

each of the first i candidates has equal probability of being the best, that's $\frac{1}{i}$

Define

$$X_i = \begin{cases} 1 & i^{th} \text{ candidate is best so far} \\ 0 & \text{otherwise.} \end{cases}$$

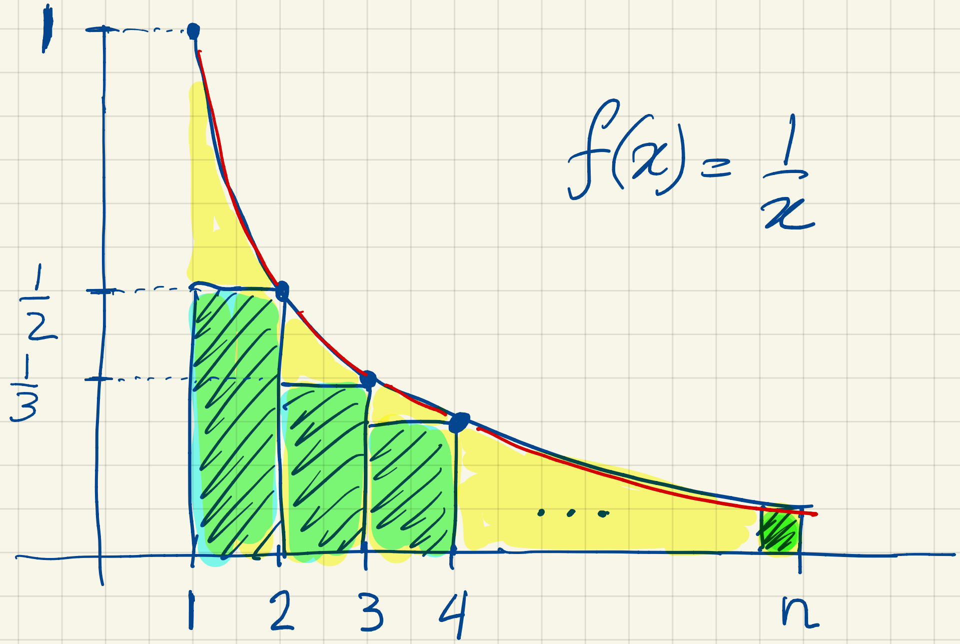
$$E[X_i] = \frac{1}{i} = 1 \cdot \frac{1}{i} + 0\left(1 - \frac{1}{i}\right)$$

let X be total number of hires, $X = \sum_{i=1}^{n} X_i$

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{i} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

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

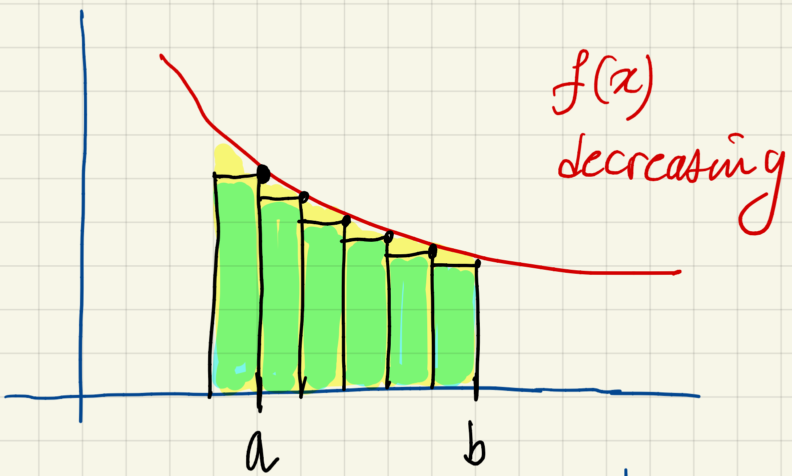
$$E[X] = \sum_{k=1}^{n} k \underbrace{P(k)}_{?}$$



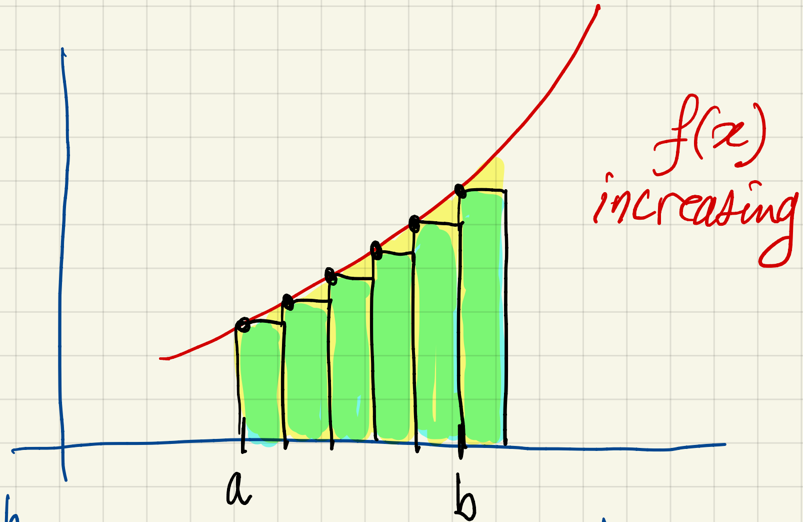
$$\sum_{i=2}^{n} \frac{1}{i} \leq \int_{1}^{n} \frac{1}{x}\,dx = \ln x \Big|_{1}^{n} = \ln n - \underbrace{\ln 1}_{0} = \ln n$$

So $E[X] = O(\log n)$

Bounding sums by integrals



$$\int_{a}^{b+1} f(x)\,dx \leqslant \sum_{i=a}^{b} f(i) \leqslant \int_{a-1}^{b} f(x)\,dx$$



$$\int_{a-1}^{b} f(x)\,dx \leqslant \sum_{i=a}^{b} f(i) \leqslant \int_{a}^{b+1} f(x)\,dx$$

Randomized Algorithm:

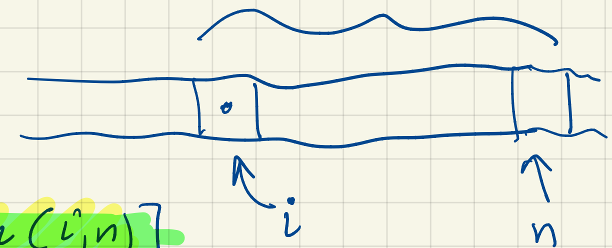
Instead of relying on the assumption about the input we enforce the random order, the alg. becomes randomized.

Hire (A, n)

for $i \leftarrow 1$ to n

do swap $A[i] \leftrightarrow A[\text{Random}(i, n)]$

⋮ as before



generates value in $\{i, i+1, \ldots, n\}$ with prob. $\frac{1}{n-i+1}$

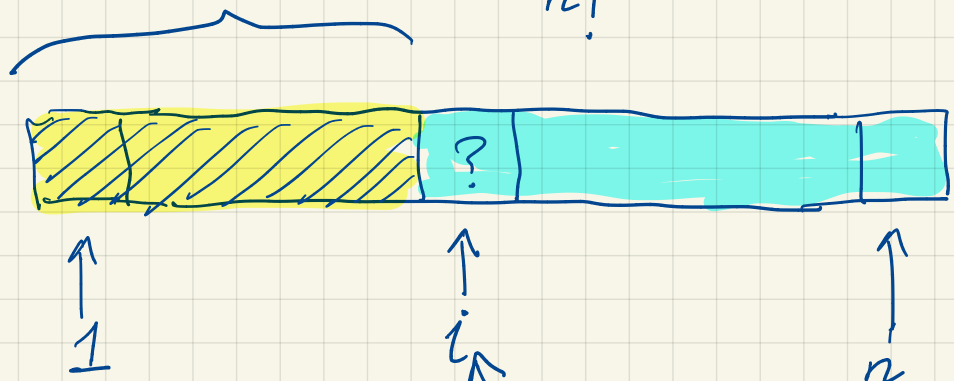
Essentially, for each position i, we assign one of the elements in $A[i \ldots n]$ randomly

(need a proof that each of the $n!$ permutations is generated with prob. $\frac{1}{n!}$)

Induction:

Assume $prob = \dfrac{(n-i+1)!}{n!}$

Base Case $i=1$: $\dfrac{(n-1+1)!}{n!} = 1$



can be any in $A[i \ldots n]$ with prob. $\dfrac{1}{n-i+1}$

$$\frac{(n-i+1)!}{n!} \times \frac{1}{n-i+1} = \frac{(n-i)!}{n!} = \frac{[n-(i+1)+1]!}{n!}$$

When $i = n+1$ (Termination)

$$prob = \frac{[n-(n+1)-1]!}{n!} = \frac{0!}{n!} = \frac{1}{n!}.$$

Back to Quicksort:

Without loss of generality, assume the elements are $\{1, 2, 3, \ldots, n\}$

Define (for $i<j$):

$$X_{ij} = \begin{cases} 1 & i \,\&\, j \text{ are compared} \\ 0 & \text{otherwise.} \end{cases}$$

i & j are compared at most once, and this happens if either i or j is the first pivot in $\{i, i+1, \ldots, j\}$ otherwise i & j go in separate ways. This has prob. $\dfrac{2}{j-i+1}$

Total number of comparisons

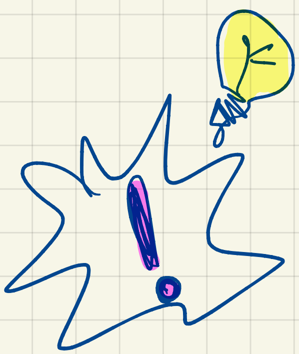
$$X = \sum_{i} \sum_{j>i} X_{ij} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$$

$$E[X] = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E[X_{ij}] = \sum_{i=1}^{n-1}\underbrace{\sum_{j=i+1}^{n}\frac{2}{j-i+1}}$$

$$\leq 2\sum_{i=1}^{n-1}\left(\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n-i+1}\right)$$

$$\leq 2\sum_{i=1}^{n-1}\left(\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right) \leq 2n\sum_{j=2}^{n}\frac{1}{j}$$

$$= O(n\log n)$$



: $$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} = \Theta(\log n)$$

n^{th} Harmonic number