NOTES TO PRESENT ON 9.30.2021

Note: What we covered on Tuesday is likely more intuitive than what we will learn today & going forward w/ PROBABILITY

$\boxed{\text{SAMPLE SPACE?}}$ ⇒ SET OF ALL POSS OUTCOMES

eg. Flip coin twice
$S = \{HH, HT, TH, T$

What's a Random Variable?
: A function X that maps OUTCOMES to $\mathbb{R}$

examples?
Flip a coin twice, get paid \$3 for each H, −\$2 for each T

X:

$HH \rightarrow \$6 \quad (3+3)$
$HT \rightarrow \$1 \quad (3-2)$
$TH \rightarrow \$1 \quad (3-2)$
$TT \rightarrow \$\ -\$4 \quad (-2-2)$

Consider a "loaded" die & a R.V. on outcomes of 1 roll mapped to value of die

X: OUTCOME | Value of die
$1 \rightarrow 1$
$2 \rightarrow 2$
$3 \rightarrow 3$
$4 \rightarrow 4$
$5 \rightarrow 5$
$6 \rightarrow 6$

Suppose the die "loaded" probability weights

1	2	3	4	5	6
1/2	1/10	1/10	1/10	1/10	1/10

What's the EXPECTED VALUE (what do we expect the value of X to be?)

$$1 \times \frac{1}{2} + 2 \times \frac{1}{10} + 3 \times \frac{1}{10} + 4 \times \frac{1}{10} + 5 \cdot \frac{1}{10} + 6 \cdot 1/10$$

$$= \frac{1}{2} + (2+3+4+5+6) \cdot \frac{1}{10} = \frac{1}{2} + \frac{20}{10} = 2.5$$

$\boxed{\text{EXPECTED VALUE}}$ is what we expect, on average, over a long time, i.e. over many TRIALS

Expected Value of a R.V. X

$$E[X] = \sum_{a \in X} a \cdot p_X(a)$$

aka. EXPECTATION of X
MEAN of X.

Two important FACTS about R.V.'s

Let X_1 and X_2 be R.V.'s
Let $X = X_1 + X_2$

e.g. X_1 could be value of die
X_2 " " 1/value of die

1. $E[X] = E[X_1] + E[X_2]$

this is, remarkably true even when X_1 & X_2 are NOT independent.

2. if c is a constant,
$E[cX] = cE[X]$

These 2 facts, together, are the LINEARITY OF EXPECTATION

What we just learned about probability will allow us to analyze randomized algorithms.

Now Ch. 5.

Ex. Problem
We want to hire an office asst. Hiring agency sends 1 candidate per day. We are cut-throat — if new candidate is better than current office asst., fire & hire.

Costs are involved: C_i to interview
C_h to hire

$C_h > C_i$

n = # of candidates interviewed (≡ # of days)
m = # of times we hire

So cost of this process is $c_i n + c_h m$.

What's WORST CASE? (most cost).

Notice we always are going to pay $c_i \cdot n$
the only thing ~~that~~ can change is # of times we hire $= m$

if $m = n$, that is we get a better candidate every day
(that is, the candidates are sent to us in increasing order of goodness)

$\Rightarrow n \cdot (c_i + c_h)$

We don't know what order we will be given the candidates
So it is natural to ask
what is the expected cost of this hiring process?

The expected cost of hiring in that manner will

result from analyzing the process when the candidates are presented in a random order. (of goodness).

An Indicator R.V. converts between PROBABILITIES and EXPECTATIONS.

If S is a Sample Space and A is an EVENT

$$X_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

X_A is an indicator R.V. for EVENT A

$$E[X_A] = 1 \cdot P(A \text{ occurs}) + 0 \cdot P(A \text{ doesn't occur})$$
$$= P(A) \quad (\text{that is probability } A \text{ occurs.})$$

In the hiring problem set up an Indicator R.V.

$$X_i = \begin{cases} 1 \text{ -- if candidate } i \text{ is hired} \\ 0 \quad \text{otherwise.} \end{cases}$$

$X \equiv$ # of times we hire (X is an R.V.)

We want to find the EXPECTED VALUE of what?

$E[X] = ?$

Write X in terms of the X_i's (the indicator R.V.'s)

$X = X_1 + X_2 + \cdots + X_n$

$$E[X] = E[X_1 + \cdots + X_n] = E\left[\sum_{i=1}^{n} X_i\right]$$

by linearity of expectation we get

$$= \sum_{i=1}^{n} E[X_i]$$

and b/c EXPECTATION of an Ind. R.V. is just PROB.

that it occurs we get

$$= \sum_{i=1}^{n} P(\text{cand. } i \text{ is hired})$$

Note: $P(\text{cand. } i \text{ is hired}) = ?$

We said the candidates come in a random order
(that is candidate i has equal chance of being best as $i-1, i-2, \ldots, 1$ among the i candidates)
therefore $P(\text{cand. } i \text{ is hired}) = \frac{1}{i}$

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

this is harmonic series.

Appdx A.7

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

$$= \sum_{k=1}^{n} \frac{1}{k}$$

$$= \ln n + O(1)$$

$$= \ln(n) + O(1)$$

That is, when we interview n candidates we hire $\ln(n)$ on average (~~when~~ they are presented in random order.

That HARMONIC SERIES is the third one to memorize

$\sum_{i=1}^{n} i$	$\sum_{i=0}^{n} x^i$	when $\lvert x \rvert < 1$ $\sum_{i=0}^{\infty} x^k$	$\sum_{i=1}^{n} \frac{1}{i}$
$= \frac{n(n+1)}{2}$	$= \frac{x^{n+1}-1}{x-1}$	$= \frac{1}{1-x}$	$= \ln(n) + O(1)$

So EXPECTED COST of Hiring, given n candidates in random order

is $\Theta(C_h \cdot (\ln(n))) + \Theta(c_i n)$

always.

If agency sends candidates in worst order get worst case cost of $\Theta(n \cdot (c_i + c_h))$

Let's use an Ind. R.V. to analyze expected Runtime of Randomized QuickSort.

Refer to Pseudocode pdf

Stopped here 9/30/202

We want to count the # of times the compare happens inside the for loop (all other stuff happens in constant time

The difference in the analysis we will do here is ~~&~~ instead of determining the # of compares ~~to~~ in ONE call, to partition & count the # of times partitions is called, we instead determine the total # of compares over all calls