DS-221 'Post Mid'

'RANDOM VARIABLES'

If we toss a six sided die and for each respective outcome we earn certain amount of money.

0	1	2	3	4	5	6
10/-	20/-	30/-	40/-	50/-	60/-	70/-

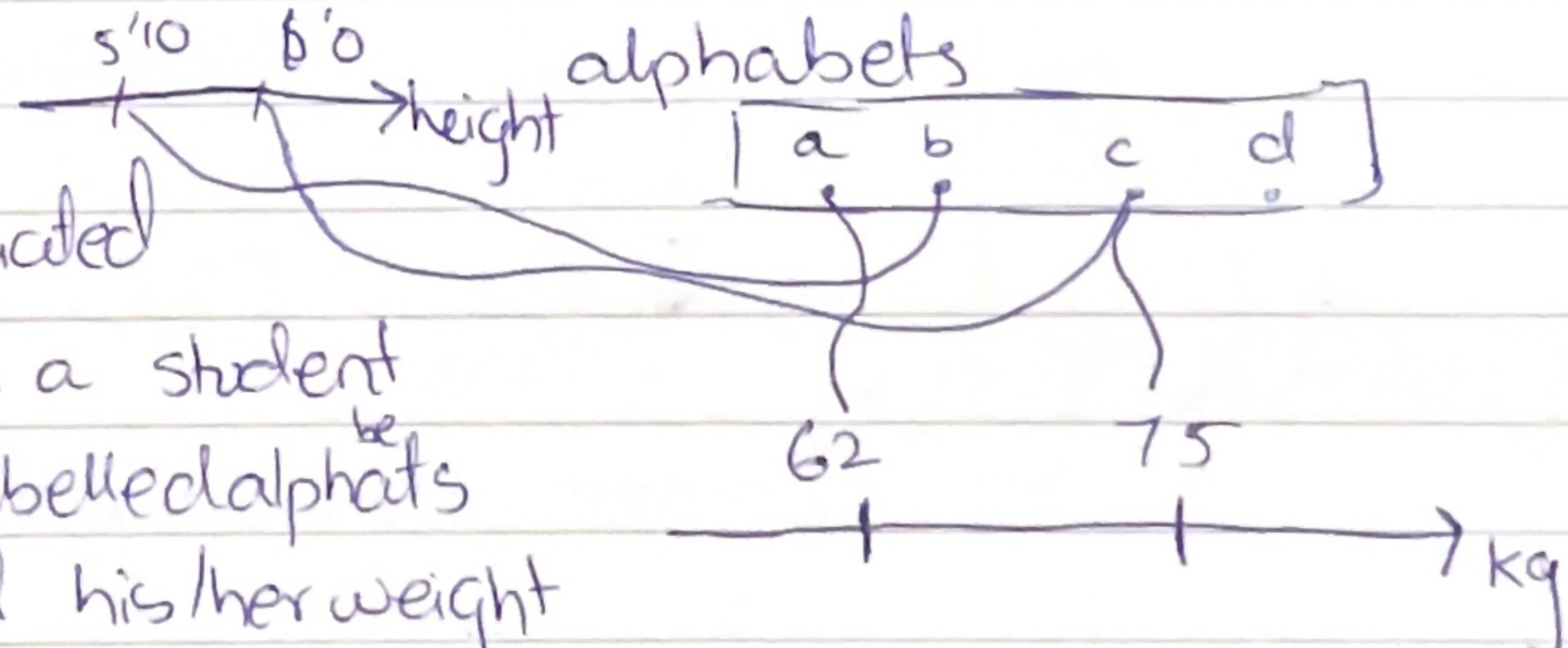
Now using probability can we answer that how much should on avg, if the die is rolled many times?

Probabilities tells about the chance of each outcome while a random variable helps to calculate the average payout; because a random variable maps to some numerical value based on a certain event / once outcome of an event is determined (that we can measure or count).

Random variable is a function that maps to some numerical value: — (a)

Consider a sample of students labelled with alphabets

(Note that probabilistic exp. can have several associated Random variable)



Now pick a student from the labelled alphabets and record his/her weight

That is how (a) is satisfied as the definition of a random variable.

$\xrightarrow{\text{student}} \boxed{\text{some funct}} \rightarrow$ student's weight.

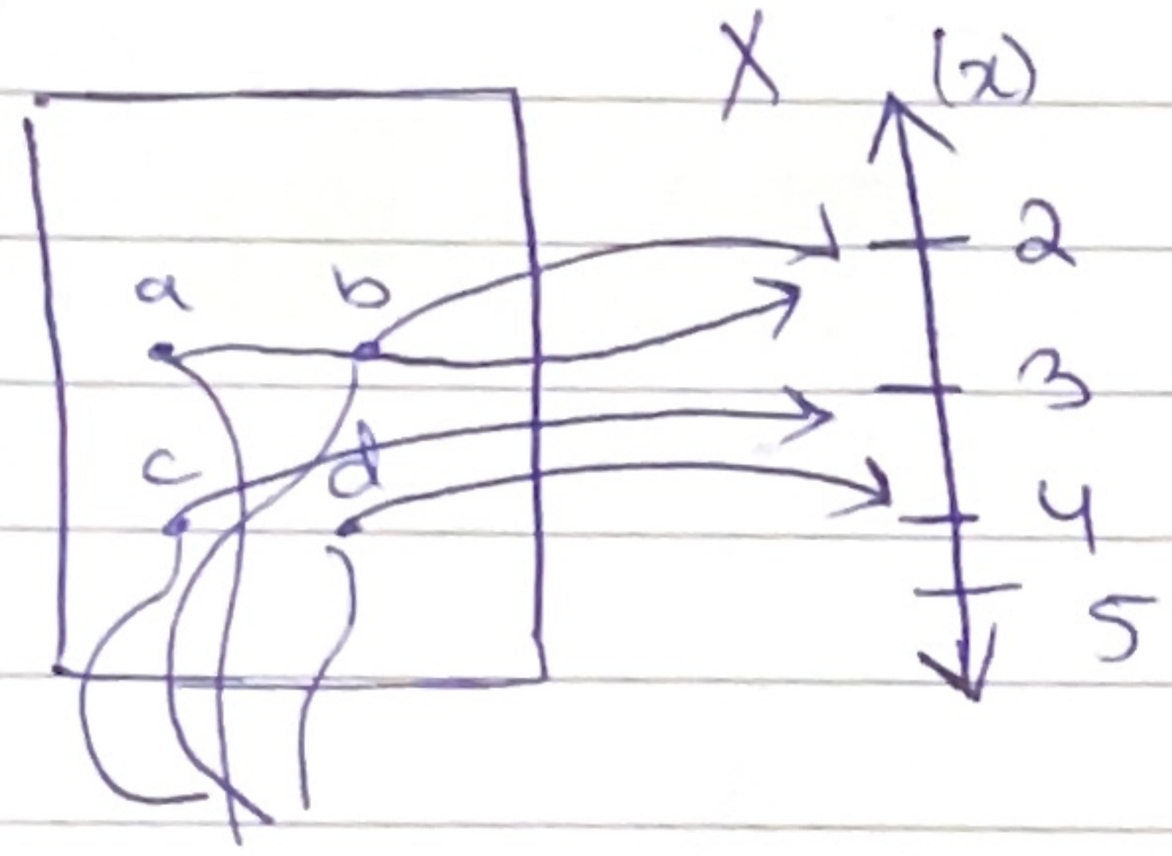
Date

Note that a new Random variable that is created based on given random variables will have indirect relation with the original dataset.

RANDOM VARABLE & PMF:-

A random variable can take different values based on the outcome of an event, e.g while tosses two coins / dies together the sum (considered a RV) can be d/f given that thar are different values possible as an outcome of the event (i.e tossing two dies).

We can define the probability of a Random Variable having a certain value by Probability Mass distribution (PMF) aka probability dist^n of X. (Random Variable).



Now X is a random variable with possible values (x) i.e 2, 3, 4 or 5.

Prob = 1/4

Now what is the probability of a random variable to have the value of 2.

$P_X(2)$ => well its depending upon two events a & b.

$\Rightarrow \frac{1}{4} + \frac{1}{4} \Rightarrow \frac{2}{4} \Rightarrow \frac{1}{2}$.

could be a possible question of identifying R.V's type.

ask gpt for examples of
→ Bernoulli R.V
→ Discrete R.V
→ Binomial R.V

Date

→ Practise plotting the PMF outcomes (i.e probabilities of a certain Random variable).

Expected Value:-

Also known as the mean (average) of the Random Variable.

↳ sum of all the random variable while each being multiplied by their possibility.
↳ If the random variables have probabilities equally likely then.

$$E(X) = \frac{1}{n}\sum x_i$$

where X is RV.

where x_i are possible values of the RV.

where $\frac{1}{n}$ → no. of samples.

Example:-

Avg no. of icecreams per customer?

→ Sum of values of RV set xlied with their prob

$E(X) = \sum x P(x)$ ⟶ I prefer this formula. (a)

$P_1(X) = \frac{225}{500}$ (Probability of 1 icecream being bought by a no. of customers.)

$P_2(X) = \frac{170}{500}$ $P_4(X) = \frac{20}{500}$ $P_4(X) = \frac{10}{500}$

$P_3(X) = \frac{55}{500}$ $P_5(X) = \frac{20}{500}$

Now the avg will be given by the formula (a).

Date

$E(x) \Rightarrow 1 * P_1(x) + 2 * P_2(x) + 3 * P_3(x) + 4 * P_4(x) + 5 * P_5(x)$
$6 * P_6(x)$

$E(X) \Rightarrow 1 \times 0.45 + 2 \times 0.34 + 3 \times 0.11 \times 4 \times 0.04 + 5 \times 0.04$
$+ 6 \times 0.61 \Rightarrow 1.94.$

Out of 200 customers, how many do you expect to buy more than 3 ice-creams?

$X > 3$

$P(X=4) \Rightarrow 0.04$
$P(X=5) \Rightarrow 0.04$
$P(X=6) \Rightarrow 0.02$

$P(X>3) \Rightarrow P(X=4) + P(X=5) + P(X=6) \Rightarrow$
$0.04 + 0.04 + 0.02 \Rightarrow 0.10.$

Now we have the expected probability of customers buying $X>3$ icecreams; its basically the 10 percent of any given set.
for 200 Customer $\Rightarrow 200 \times 0.10$
$\Rightarrow 20$

The Random Variable's values can be represented by functions as well, in that form the formula becomes.

$E(x) = \sum g(x) P_x(x)$

** remember $E(X^2)$

Date $\Rightarrow \sum a^2 P(x)$

underroot of Variance is std. deviation.

Consider the following example.

$$P_X(x) = \begin{cases} 1/4 & x = 1,2,3,4 \\ 0 & \text{otherwise} \end{cases}$$

$$Y = g(x) = \begin{cases} 10.5X - 0.5X^2 & 1 \leq X \leq 5 \\ 50 & 6 \leq X \leq 10 \end{cases}$$

Now the probability of Random variable ranging from $1 \rightarrow 4$ are given.

however the random variable's are further represent by the function $g(x)$.

$$\therefore \; E(X) = g(x)P_x(x)$$

$$\Rightarrow (10.5X - 0.5X^2)P_1(X) + (10.5X - 0.5X^2)P_2(X) + (10.5 - 0.5X^2)P_3(X) + (10.5 - 0.5X^2)P_4(X).$$

$$\Rightarrow \frac{1}{4}(10.5(1) - 0.5(1)^2) + \frac{1}{4}(10.5(2) - 0.5(2)^2) + \frac{1}{4}(10.5(3) - 0.5(3)^2) + \frac{1}{4}(10.5(4) - 0.5(4)^2)$$

Variance:-

Spread of / around the expectation is also important.

Some we may have same expectation/avg values for entirely different datasets/random variable thus variance would help to further analyze our take on information.

$$\sigma^2 = Var(X) = E(X - EX)^2 \Rightarrow \sum_x (X - \mu)^2 P(x)$$

$$\text{or } \sigma^2 = Var(X) = [E(X^2) - E(X)^2]$$