Till now

- Descriptive statistics
- Probability
  - Probability distribution.

Agenda

- Probability distribution.

random variable

- Discrete → Probability mass function
  1. Discrete Uniform Dist.
  2. Bernoulli
  3. Binomial
  4. Poisson dist
  5. Geometric distribution.
- Continuous → Prob density function.
  1. Normal distribution
  2. Standard Normal distribution
  3. Log-Normal distribution
  4. Chi-square distr.
  5. F distribution
  6. Exponential distribution
  7. Continuous uniform distribution
  8. t-distribution.

Uniform < Discrete / Continuous.

① Discrete Uniform distribution ($U(a,b)$)

→ Discrete

→ A uniform distribution refers to a type of prob distribution in which outcomes are equally likely.

→ In discrete uniform distr the outcomes are discrete in nature.

→ The outcomes are independent.

eg. rolling a dice.

eg. tossing a coin

eg. Picking up card from well shuffled deck of card.

dice → 1, 2, 3, 4, 5, 6

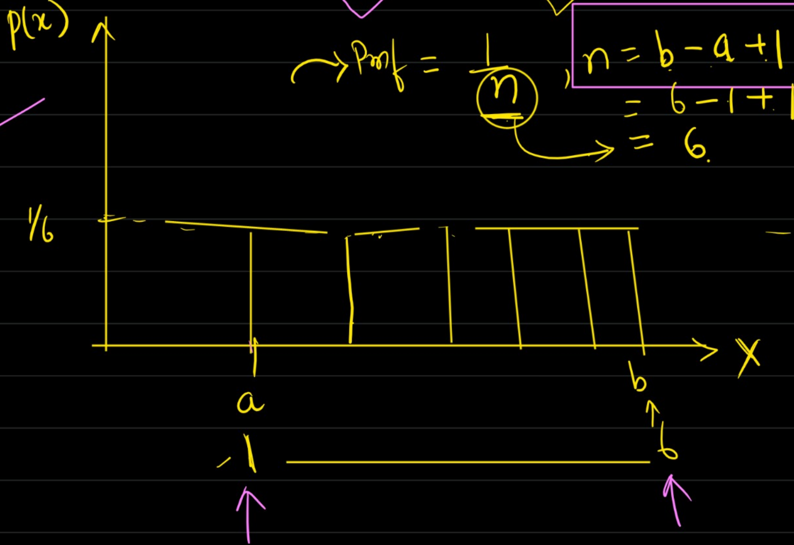
✓ 1st — 3 → 1/6

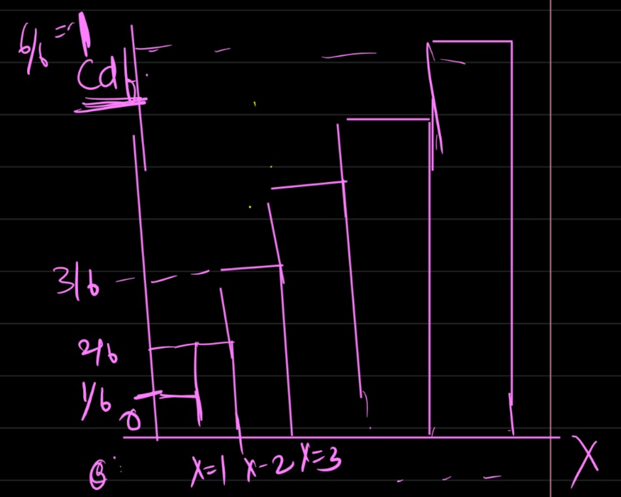
✓ 2nd — → 1/6

pmf

$p(x)$

$\rightarrow \text{Pmf} = \frac{1}{n}$, $\boxed{n = b - a + 1}$
$= 6 - 1 + 1$
$= 6$





throwing a dice → 1, 2, 3, 4, 5, 6 —— X —— X →

✓ mean of discrete Uniform distribution $= \frac{a+b}{2}$ ✓

✓ variance of " " " $= \frac{n^2-1}{12}$ ✓

mean = sum of nos divided by total nos

⇓

Expected value → In the long run avg value of repetitions of experiments.

$\sum_{i=1}^{6} x \cdot p(x)$

	X	p(x)	X · p(x)
0.17	✓ 1	1/6	1×1/6 = 0.167 × 1
0.33	2	1/6	2×1/6 = 0.167 × 2
0.50	3	1/6	3×1/6 = 0.167 × 3
0.67	4	1/6	4×1/6 = 0.167 × 4
0.83	5	1/6	5×1/6 = 0.167 × 5
1	6	1/6	6×1/6 = 0.167 × 6

$\sum_{i=1}^{n} x \cdot p(x) \Rightarrow 3.5$

→ For U.D → 1, 2, 3, 4, 5, 6 (dice)

$\text{Mean} = \frac{1+2+3+4+5+6}{6} = 3.5$

mean & EV

$$E(x)/\mu = \sum_{i=1}^{n} x_i \, p(x_i)$$

$$Var(x), \; U(a,b) = \frac{n^2-1}{12}$$

$$\boxed{Var(x) = E(x^2) - (E(x))^2}$$

$$E(X) \rightarrow \sum_{i=1}^{N} x \, p(x)$$

$$= \sum_{i=1}^{N} x \cdot \frac{1}{N} = \frac{1}{N}\left(\underbrace{1+2+3+\cdots N}_{\text{Sum of N natural no.}}\right)$$

$$\left(N \cdot \frac{(N+1)}{2}\right)$$

$$= \frac{1}{\cancel{N}}\left(\frac{\cancel{N}(N+1)}{2}\right) = \frac{N+1}{2}$$

$$E(x^2) = \sum_{i=1}^{N} x^2 \cdot p(x)$$

$$= \sum_{i=1}^{N} x^2 \cdot \frac{1}{N} \Rightarrow \frac{1}{N}\left(1^2 + 2^2 + 3^2 + \cdots N^2\right)$$

$$= \frac{1}{\cancel{N}}\left(\frac{\cancel{N}(N+1)(2N+1)}{6}\right)$$

$$= \frac{(N+1)(2n+1)}{6}$$

$$Var(x) = E(x^2) - (E(x))^2$$

$$= \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2$$

$$= \frac{2N^2+3N+1}{6} - \frac{N^2+2N+1}{4}$$

$$Var(x) = \frac{N^2-1}{12}$$

$$\sigma = \sqrt{\frac{N^2-1}{12}}$$

Ai describe var

$$Var(x) = \frac{n^2-1}{12} = \frac{6^2-1}{12} = \frac{36-1}{12} \Rightarrow \frac{35}{12} = 3$$

Pre-requisite

* Sum of first 'n' natural no $= \frac{N(N+1)}{2}$

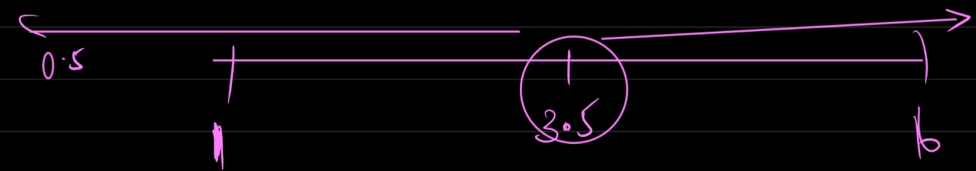
$1, 2, 3, 4, 5, 6, 7, 8, 9, 10$
$= 55$

$\frac{10(10+1)}{2} = \frac{10 \times 11}{2}$

* Sum of first N square no $= \sum x^2$

$1^2+2^2+3^2$

$= \frac{N(N+1)(2N+1)}{6}$



* Continuous Uniform distribution

* That has infinite no of values in a specified range.

→ r.v is continuous.

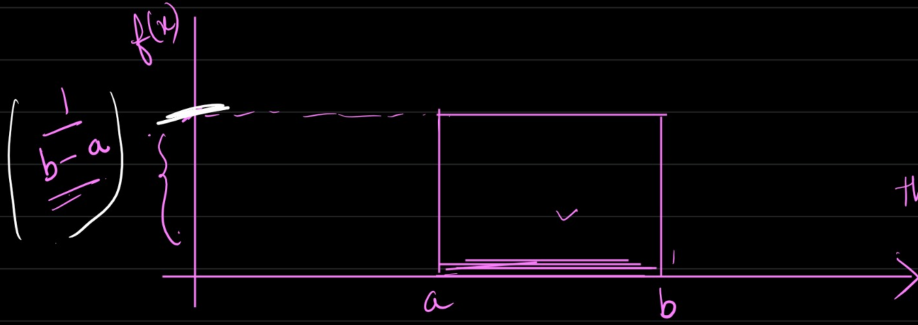
→ rectangular distⁿ.

— A perfect random no generator.
— Prob of guessing exact time at any moment
— waiting time at a bus stop.
— Temp variation in a day.

Notation :- $U(a,b)$

Parameter $-\infty < a < b < \infty$

pdf



the max prob = 1

Area under curve
↳ Probability value = 1

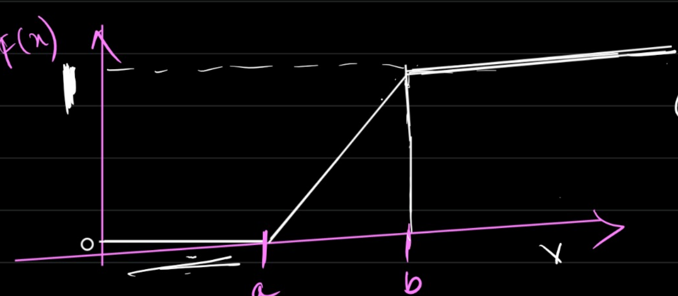
Total area in pdf gives you the prob.

Area of rectangle = $b-a \times f(x)$

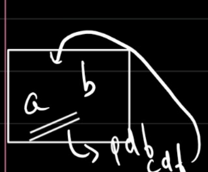
$$1 = b-a \times f(x)$$

$$\boxed{f(x) = \frac{1}{b-a}}$$

cdf →



$$\text{cdf} = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$$

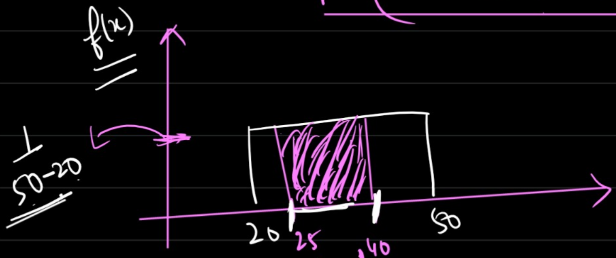


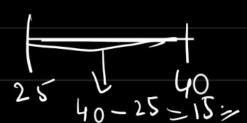
The mean of Cont Uniform dist $= \frac{1}{2}(a+b)$

" " " " " $= \frac{1}{12}(b-a)^2$

Q. The number of items sold at a shop daily is Uniformly distributed with max and min item sold 50 & 20 respectively.
What is prob of daily sales to fall between 25 & 40 ??

$\rightarrow$ $P(25 \le x \le 40) = ??$





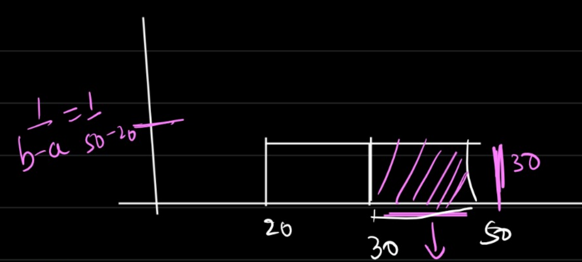
$\frac{1}{b-a} \times 15$

$= \frac{1}{30} \times 15 = 0.5$

$P(X \geq 30)$

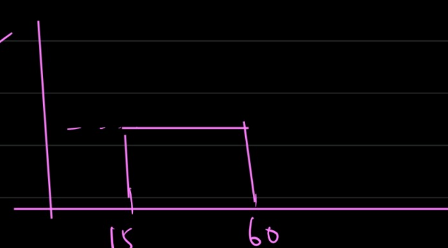
$\Downarrow$

$20 \times \frac{1}{30} = 0.66$



Q. The amount of time for Pizza delivery is Uniformly distributed b/w 15 & 60 mins. What is standared deviation of the amount of time it takes for Pizza to be delivered?

$\Rightarrow$



$var(x) = \frac{1}{12}(b-a)^2$

$= \frac{1}{12}(60-15)^2 = \frac{45^2}{12} = 168.7$ Cm.

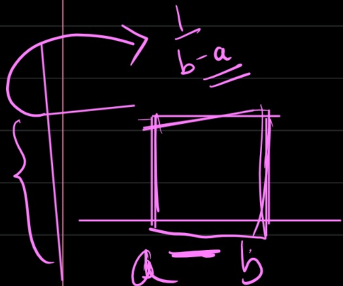
$\sigma = \sqrt{var} = \sqrt{168.75} \approx 13$ mins

* for a continuous random var prob density fn $(f x)$.

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

$\sum x \cdot p(x)$ → discrete

Continuous



$$= \int_a^b x f(x)\,dx$$

$$\int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a}\int_a^b x\,dx.$$

$$\boxed{\int x^1\,dx = \frac{x^{1+1}}{1+1} = \frac{x^2}{2}}$$

$$= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

$$\int x^2\,dx = \frac{x^3}{3}$$

$$= \frac{1}{2(b-a)}\left[x^2\right]_a^b$$

$$= \frac{1}{2(b-a)}\left(b^2 - a^2\right)$$

$b^2 - a^2 = (b-a)(b+a)$

$$= \frac{1}{2\cancel{(b-a)}}(b+a)\cancel{(b-a)} = \frac{a+b}{2}$$

$$\text{var}(x) = E(x^2) - (E(x))^2$$

$E(x) \to \frac{a+b}{2}$

$$E(x^2) = \int_a^b x^2 f(x)\,dx = \int_a^b x^2 \cdot \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a}\left[\frac{x^{2+1}}{3}\right]_a^b$$

$$= \frac{1}{3}\cdot\frac{1}{b-a}\left[x^3\right]_a^b$$

$$= \frac{1}{3(b-a)}\left(b^3 - a^3\right)$$

$b^3 - a^3 = (b-a)(b^2+ab+a^2)$

$$= \frac{1}{3\cancel{(b-a)}}\cancel{(b-a)}(b^2+ab+b^2)$$

$$= \frac{b^2+ab+a^2}{3}$$

$$E(X^2) - (E(X))^2$$

$$\left(\frac{b^2+ab+a^2}{3}\right) - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{b^2+ab+a^2}{3} - \frac{a^2+b^2+2ab}{4}$$

$$\frac{4b^2+4ab+4a^2-3a^2-6ab-3b^2}{12} = \frac{b^2-2ab+a^2}{12}$$

$$= \frac{(b-a)^2}{12}$$

distribution

pmf → Discrete UD

pdf → continuous UD

two possible outcome → binary

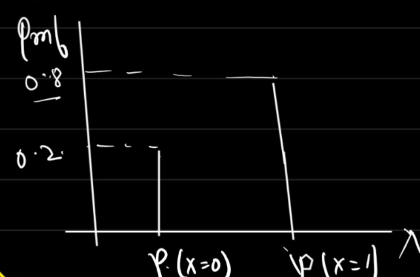
② Bernoulli distribution:

A discrete prob distribution of a random varianble which takes only two possible outcomes typically labelled as success (1) and failure (0).

* To model a problem statement with only two possible outcomes

eg. Toss a coin. (Head or tail)

$$\begin{cases} P(X=0) = 0.5 = p \\ P(X=1) = 1-0.5 = q \quad (1-p) \end{cases}$$



$$\text{pmf} \Rightarrow P(X=k) = \begin{cases} p & \text{if } k=1 \\ 1-p & \text{if } k=0 \end{cases}$$

$$\boxed{P(X=k) = p^k(1-p)^{1-k}}$$

① if $k=1$

$$p^1(1-p)^{1-1}$$

$$p^1(1-p)^0$$

$$p$$

② if $k = 0$

$$p^0 (1-p)^{1-0} = 1-p$$

* Conditions of Bernoulli disth

① No of trial should be finite.

② Each trial should be independent

③ only two possible outcome < 1, 0

④ Prob of each output should be same in every trial.

Examples — tossing a coin

→ prob of getting 3 and not 3 while throwing a dice.

→ pass | fail, fraud | not F

Classification — diabetes | not diabetes

transaction amt	Count of tran	loan	Salary	Fraud / Not fraud
1500	2	5	3	1
—	—	—	—	0
—	—	—	—	0
				1
				1
⋮				0
				0
				1

← Bernoulli dist

# Student	marks in internal exam	# test paper solv	pass / fail

* Mean and Variance of Bernoulli distribution

Mean: p

Variance: $p(1-p)$

① Mean $E(X) = \sum_{i=1}^{k} x \cdot p(x)$

$= x=1 \;(0.6) + x=0 \;(0.4)$

$= 1 \times 0.6 + 0 \times 0.4 = 0.6 \rightarrow p$

② Variance $\quad E(x^2) - E(x)^2$

$$E(x)^2 \rightarrow (p)^2$$

$$E(x)^2 = \sum x^2 \, p(x) \rightarrow (1-p)$$

$$= 1^2 \times p + 0 \cdot (1-p)$$

$$= \underline{\underline{p}}$$

$$E(x^2) - (Ex)^2$$

$$p - (p)^2 = p(1-p)$$

Proved

* Binomial distribution

Bi $\longrightarrow$ 2 $\longrightarrow$ two outcomes.

* Binomial distribution is 'n' Bernoulli trial

$$pmf = {}^{n}C_{k} \, \underline{p^{k} (1-p)^{n-k}}$$

For bernoulli

$$\underline{\underline{p^{k} (1-p)^{1-k}}}$$

$$\binom{n}{k} = \frac{\lfloor n}{\lfloor n-k \; \lfloor k} = \frac{n!}{(n-k)! \; k!}$$

$$ {}^{5}C_{3} = \frac{\lfloor 5}{\lfloor 3 \; \lfloor 5-3} = \frac{5!}{3! \; (5-3)!}$$

$$\lfloor 5 = 5! = 5 \times 4 \times 3 \times 2 \times 1$$

$$\lfloor 3 = 3! = 3 \times 2 \times 1$$

$$\frac{\lfloor 5}{\lfloor 3 \; \lfloor 2} = \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1}$$

1st toss 2nd toss 3rd toss 4th toss 5th toss ------ 10th toss.

→ what is prob of getting 'K' heads out of n trials.

$n \rightarrow$ total no of trials

K – the no of events that you are interested in.

Example: with 3 tosses what is probability of getting exactly 2 heads ??

$^nC_k \, p^k (1-p)^{n-k}$

$n = 3$, $K = 2$, $p = 1/2$

$$p(x=2) = {}^nC_k \, p^k (1-p)^{n-k}$$

$$= {}^3C_2 \left(\frac{1}{2}\right)^2 \left(1-\frac{1}{2}\right)^{3-2} = \frac{\lfloor 3}{\lfloor 2 \lfloor 3-2} (0.5)^2 (0.5)$$

$$= \frac{3 \times \cancel{2}}{\cancel{2} \times 1} (0.5)^3 = 3 \times (0.5)^3$$

$$= 3 \times 0.125$$

$$= 0.375$$

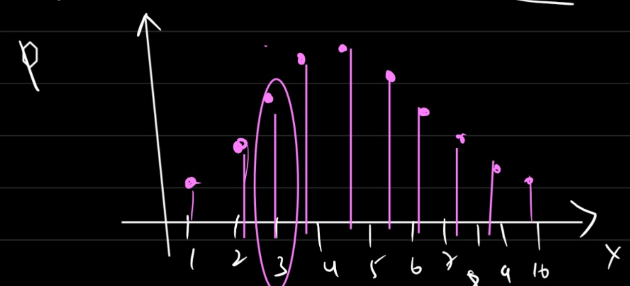
Q. When you toss a coin 10 times, what is prob that you will get head exactly 3 times.

→ $n = 10$, $k = 3$ $\quad K(x=3) = {}^{10}C_3 \cdot (0.5)^3 (0.5)^{10-3}$

$$= \frac{\lfloor 10}{\lfloor 7 \lfloor 3} \cdot (0.5)^{10} = \frac{10 \times 9 \times 8 \times \cancel{\lfloor 7}}{\cancel{\lfloor 7} \times 3 \times 2} \cdot (0.5)^{10}$$

$^{10}C_3$

pmf



$\boxed{\begin{array}{l} \text{mean} = np. \\ \text{variance} = np(1-p) \\ \qquad npq. \end{array}}$

* Geometric distribution

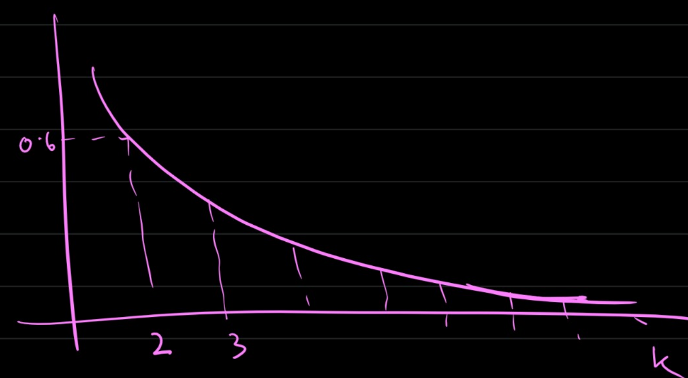
It represents the probability of getting first Success after having a consecutive no of failure (n no of trials)

1st toss   2nd toss   3rd toss   4th to - - - .

(H)

T   T   T   H

$$\text{pmf} \Rightarrow P(X=x) = (1-p)^{x-1} \cdot p$$



mean $= 1/p$

Variance $\rightarrow \frac{1-p}{p^2}$

* Poisson distribution

→ The Poisson distribution is a discrete prob distn that describes the no of events that occur within a fixed interval of time or space given a known average rate of occurrence.

* No of events occurring in a fixed time interval.

eg. No of calls recieved by a customer care every hour.

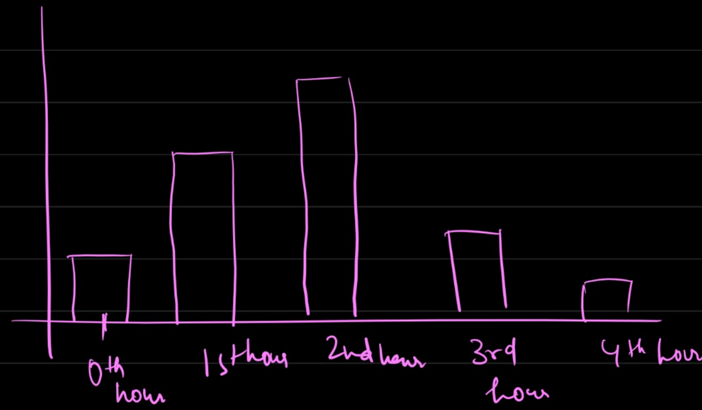
Expected no of calls every hour is $\lambda = 100$

eg. No of people going to temple

eg No of people going to hospital / airport every hour.

eg. No of emails every hour

Pmf.



$$pmf = P(X=x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!}$$

λ - avg rate of events every interval.

e - 2.718

$\lambda = 10$

Prob of a person visiting at 5th hour.

$$P(X=5) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} = \frac{(2.718)^{-10} \cdot 10^{5}}{5!}$$

Q. The average no of customers entering a store in an hour is 5. What is prob exactly 3 customers will enter the store next hour.

$\lambda = 5$

$P(X=3)$

$$\frac{e^{-\lambda} \cdot \lambda^{x}}{x!} = \frac{e^{-5} \cdot 5^{3}}{3!}$$

$$= \frac{(2.715)^{-5} \times 125}{6}$$

$$= 0.00674 \times \frac{125}{6} \approx 0.14$$

Mean — $\lambda \times t$ → time
Varian — $\lambda \times t$

Counts ⇒ (10) person every hour — Avg rate in general

→ 0 – 10

10 → 1 hour → 0 – 10 → Spread

a ——————————————— b

avg 50 km/h

in 1 hour — Avg × time
= 50 × 1
= 50 km

2 hr = 50 × 2
= 100 km