happy holi :)

Agenda

→ Exponential distribution

→ Log-Normal distribution

let's start @ 9:05 pm

Rate of Scoring goals

2.5 goals / 90 min → rate.

λ or μ 'mu'

Matchlevel

poisson.pmf ($k=2$, $\mu = 2.5$)

45 min → $\boxed{\lambda = 1.25}$

30 min → $\boxed{\lambda = 0.8}$

$P(k=5)$ $k=5$
$\lambda = 2.5$

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

240 messages / hour → Average.

① What is the no. of messages in 30 sec?

3600 sec → 240

$1 \rightarrow \frac{240}{3600}$

$30 \rightarrow \frac{240 \times 30}{3600} = 2$

$\lambda = 2$

② What is the prob. of 1 msg in 30 sec.

poisson.pmf(k=1, mu=2) ✓

3 No messages in 10 sec. ?

↳ λ for 10 sec. ✓ $= \frac{2}{3}$

↳ poisson.pmf(k=0, mu= $\frac{2}{3}$)

$\boxed{240 \text{ msg} / \text{hr}}$ per 3600 sec.

① What is the avg. time to wait b/w 2 messages?

240msg → 3600

0 2 4 6 10 → 5 msg → 10 sec.

$\frac{10}{5} = 2$ sec.

$\frac{3600}{240} = 15$ sec

② What is avg. # of msg per sec.?

3600 s → 240 msg

1 → $\left(\frac{240}{3600}\right) = \left(\frac{1}{15}\right)$

$= 0.067$

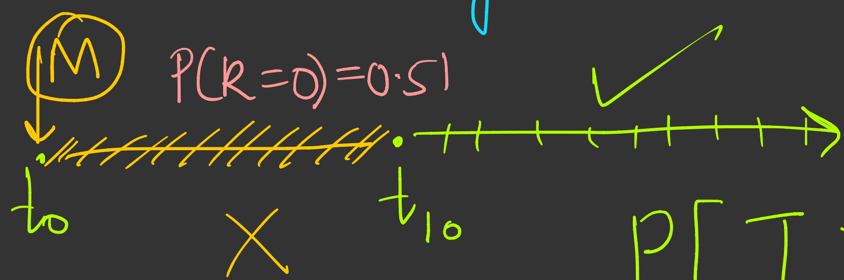
λ_1

(3) What is the prob. of no msg. in 10 sec.

$$\lambda_{10} = 10\cdot\left(\frac{1}{15}\right) = 10(\lambda_1)$$

poisson.pmf(k=0, mu=10/15) = 0.51 ✓

(4) What is the prob. of waiting more than 10 sec. for the next message?



$$P[T>10] = ?$$

- Count → Avg. count in fixed time interval
  (discreate)
- time → Avg. time for next msg.
  (Contincous)
- Let 't' denote time for next msg.
  waiting time
  Random Variable → $\sim$ dbn ? ✓

3rd

$k = 0$

$\lambda_{10} = 10 \cdot \left(\frac{1}{15}\right) \Rightarrow P(X=k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$

$$P(X=0) = \frac{(10/15)^0 \cdot e^{-10(1/15)}}{0!}$$

$$\boxed{P(X=0) = e^{-10/15}} = P[T > 10]$$

Prob. of getting my next msg. after 10 sec.

$\lambda_1 = \left(\frac{1}{15}\right) \longrightarrow$ rate/sec.

$$P(T>10) = e^{-10\cdot\left(\frac{1}{15}\right)}$$

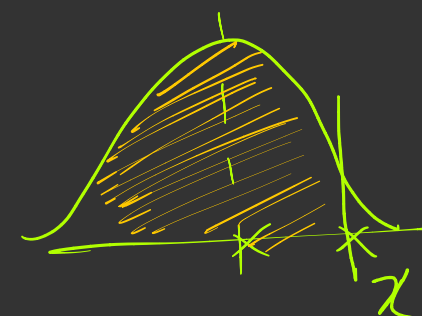
$$\boxed{P(T>10) = e^{-10\cdot\lambda_1}}$$ ✓

$\Rightarrow$ For any specific 'x'

$$\boxed{P(T>x) = e^{-x\cdot\lambda_1}}$$ ✓

(5) What is the prob. of waiting less than 10 sec. for next msg?

$$P(T \le 10) = 1 - e^{-10 \cdot \lambda_1}$$



$$\boxed{P(T \le x) = 1 - e^{-x \cdot \lambda_1}}$$

expon.cdf(   )

$P(X \le x)$

$\downarrow$

norm.cdf.

CDF of Exponential distribution ✓

T ~ Exp. dbⁿ

Scale = 5
$\lambda = 1/5$

Scale → Avg. time b/w 2 Events

scale = 15 sec

Dev → debug

"give me 5 min"

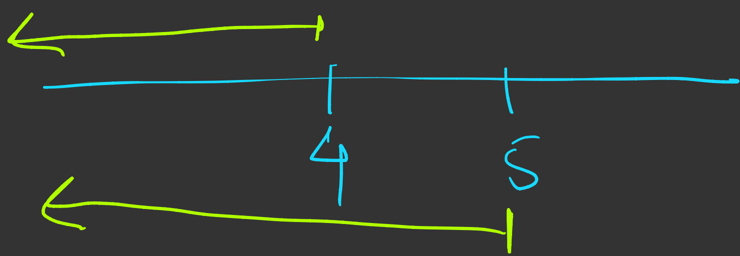
$\boxed{\text{Scale} = 5 \text{ min}}$

avg. time to debug → 5 min ~ Exponential

Q1 Prob. of debug in 4 to 5 min?

$$P[\underset{x}{4} < T < \underset{x}{5}] = \text{cdf}(x=5, \text{scale}=5) - \text{cdf}(x=4, \text{scale}=5)$$

$$= 0.08 \checkmark$$



Q2.) Prob. of debug more than 6 min ✓

$P(T > 6 \text{ min}) = 1 - \text{expon.cdf}$

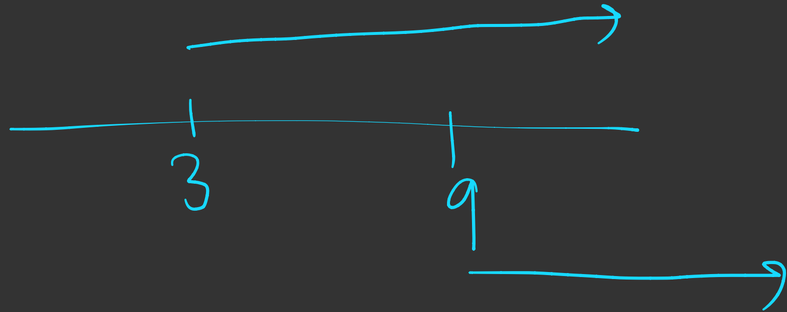
$= 0.301$ ✓ $(x=6, \text{scale}=5)$

Q3.) given that he has already spent 3 min. What is the prob. that he will take more than 9 min. to debug?

$\Rightarrow P[T > 9 \mid T > 3]$

$= \dfrac{P[(T>9) \cap (T>3)]}{P(T>3)}$

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$



$$P(T>9)$$

$$P(T>9 \mid T>3) = \frac{P(T>9)}{P(T>3)}$$

$$= \frac{1 - \text{expon.cdf}(x=9, \text{scale}=5)}{1 - \text{expon.cdf}(x=3, \text{scale}=5)}$$

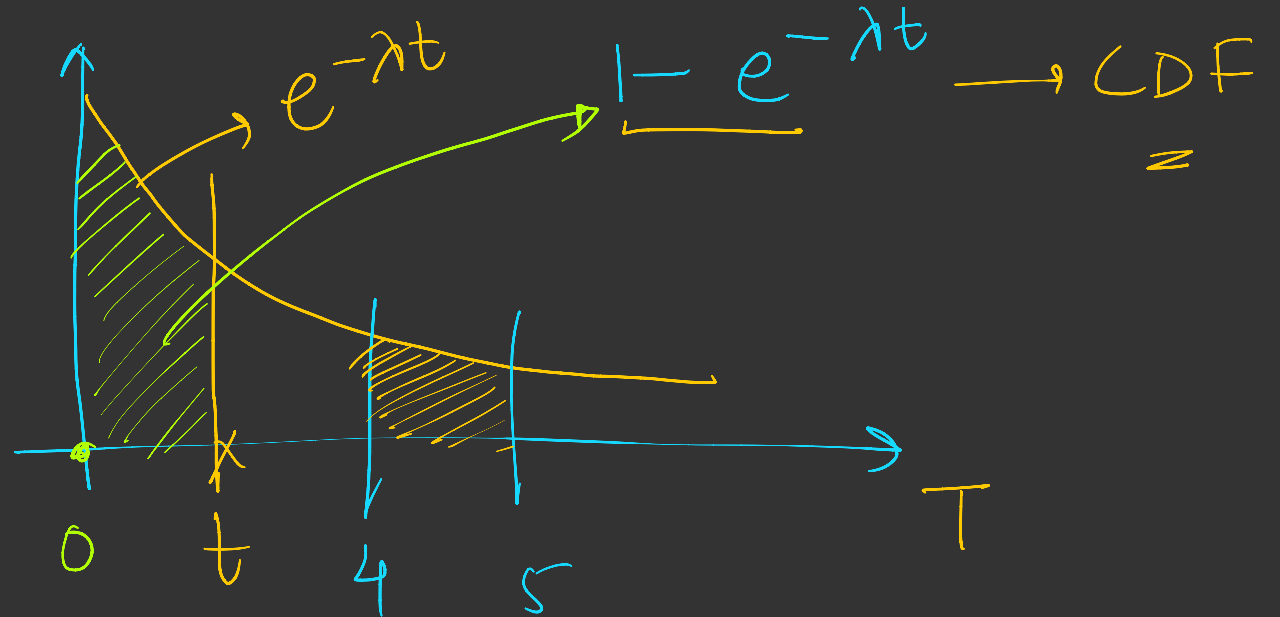
$$= 0.301$$

$$P(T>3) \longrightarrow P(T>9) \quad \checkmark \; = 0.301$$

$$P(T>6) \quad \checkmark \; = 0.301$$

$$P[T>9 \mid T>3] = P[T>6]$$

memoryless property.

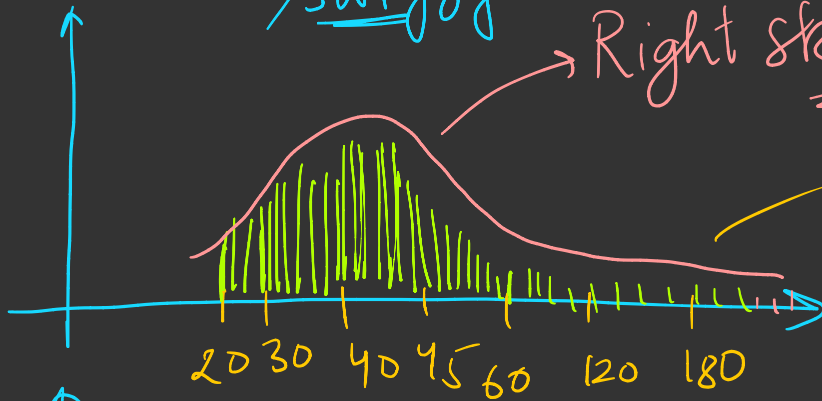


$e^{-\lambda t}$
$1 - e^{-\lambda t}$
CDF
x
T
0
t
4
5

log-normal :

Swiggy → delivery time

Right skewed

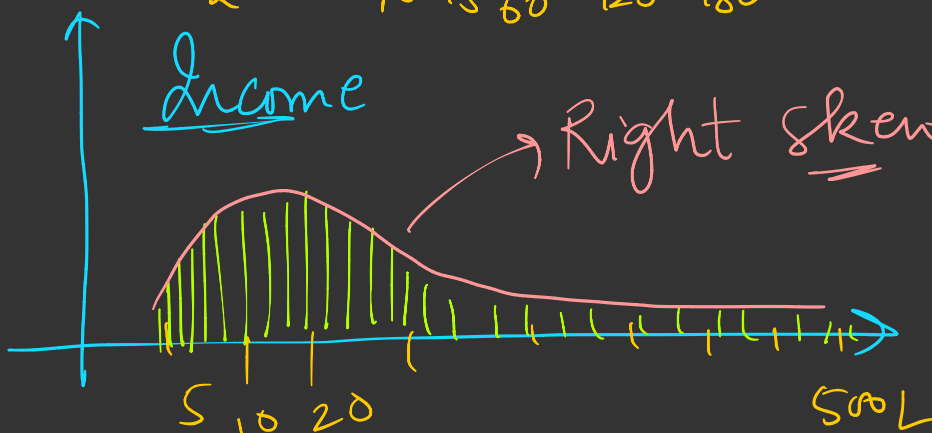


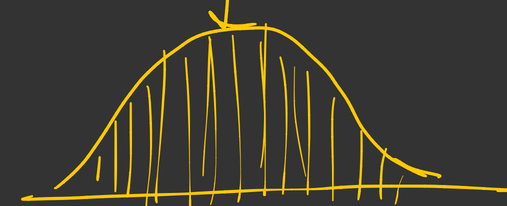
→ log-normal

x → log(x)

Income

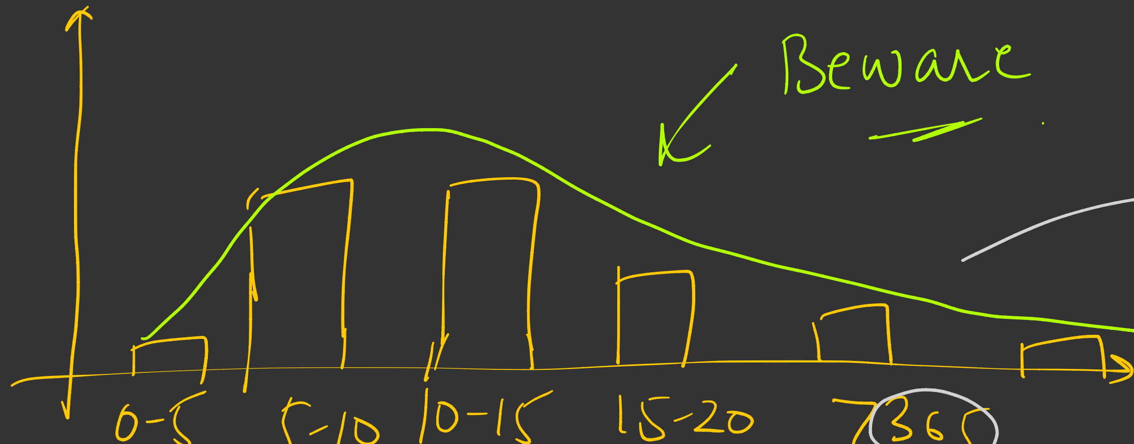
Right skew

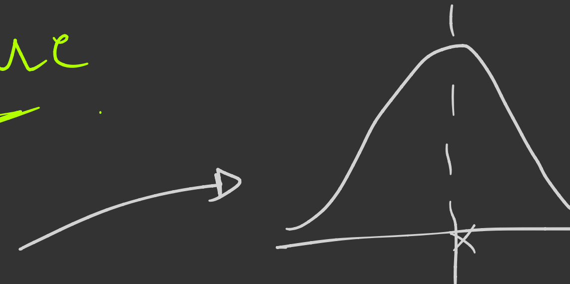




hospitalization days

Beware





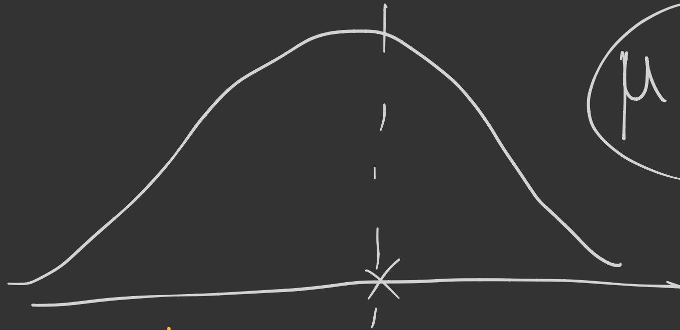
data

$$\begin{bmatrix} 5 \\ 6 \\ 20 \\ \vdots \\ 3000 \end{bmatrix}$$

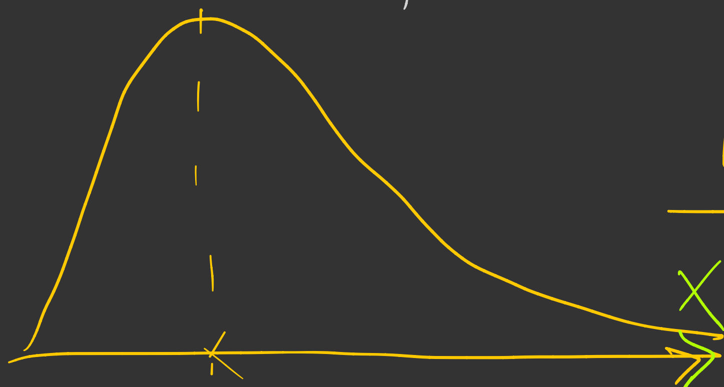
$\log_{10} 10 = 1$

$\log_{10} 100 = 2$

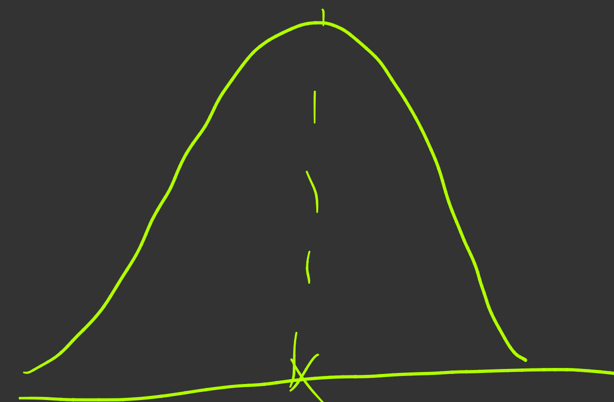
$\log_{10} (10)^{16} = 16$



μ, σ ✓



$$\xrightarrow{\log(X)}$$



(μ, σ^2)

$$\Rightarrow \boxed{E(X) = e^{\left(\mu + \sigma^2/2\right)}}$$

$$\boxed{\mathrm{Var}[X] = \left(e^{\sigma^2} - 1\right) \cdot e^{(2\mu + \sigma^2)}}$$

Income → Income_log → model algo

Income	
5	[2, 5000L]
6	
120	
300	
18	
20	

Income_log: [2 to 7]

model is trained → good results

~~500L~~

$\log(500L) \approx$ very small (acceptible range)

Standard transformation

$$\begin{bmatrix} -5 \\ \vdots \\ -10 \\ \vdots \\ 10 \\ 20 \end{bmatrix} \rightarrow x_i$$

$$\frac{\bar{x}}{\sigma}$$

$$\underbrace{\left(\frac{x_i - \bar{x}}{\sigma}\right)}$$

$$\downarrow$$

$$\left(\frac{-5 - \bar{x}}{\sigma}\right) \rightarrow \underbrace{\text{z score}}$$

New Salary $\rightarrow$ 200000 $\rightarrow$ z score.