Exponential Distribution

Example: Q. You receive 240 messages per hour on average - assume Poisson distributed. Rate of messages arriving per second is $\frac{1}{15}$.

Q1) What is the probability of having no message in 10 seconds? → $P(X=0)$

↓ message

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

$k \Rightarrow 0$

$\lambda \Rightarrow \frac{1}{15} \times 10$

$$\frac{\left(\frac{10}{15}\right)^{0} * e^{-10/15}}{0!}$$

$$\Rightarrow e^{-\frac{10}{15}}$$

Assume λ is denoted as rate per sec $\lambda = \frac{1}{15}$, the

$$P(X=0) = e^{-10\lambda}$$

Q2. What is the probability of waiting for more than 10 seconds for the next message?

$P(T>10)$ $e^{-10\lambda}$

↓

Continuous

Q3. What is the probability of waiting less than or equal to 10 seconds?

$$P(T \le 10) \Rightarrow 1 - e^{-10\lambda}$$

rate

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

Scale $\beta = 1/\lambda$ (inversely proportional)

$$P(T \le x) = 1 - e^{-\frac{x}{\beta}}$$

$$f(x : \lambda) = 1 - e^{-x\lambda} \text{ or } 1 - e^{-x/\beta}$$

* C.D.F

$$\boxed{1 - e^{-x/\beta}} \longrightarrow 1 - e^{-\lambda x}$$

* P.D.F

$$\frac{1}{\beta} * e^{-x/\beta} \longrightarrow \lambda e^{-\lambda x}$$

Poisson Distribution vs Exponential Distribution

Poisson Distribution: $X \Rightarrow$ discrete

- **Use Case:** Models the number of events in a fixed interval of time or space.
- **Example Question:**
  - "How many customers will enter a store in the next hour?"
  - "How many messages will you receive in next 15 mins?"
  - "How many calls can the call center expect in the next 30 minutes?"
- **Parameter:** Rate (λ) represents the average number of events in the specified interval.

Exponential Distribution: $\Rightarrow$ (T) continuous

- **Use Case:** Models the time between consecutive events.
- **Example Question:**
  - "How long do I have to wait for the next message?"
  - "On average, how much time will a customer spend waiting for service in a queue?"
  - "How long, on average, will passengers wait between consecutive bus arrivals?"
- **Parameter:** Scale represents the average time between events. It's the reciprocal of the rate.

Questions

$\lambda = \frac{}{\text{Per unit Time}}$ $\beta = \frac{}{\text{Per-event}}$

```
You are working as a data engineer who has to resolve any bugs/
failures of machine learning models in production.
The time taken to debug is exponentially distributed with mean of 5 minutes
```

1 event →

1 → 5 mins

How many Bugs in 1 min ⇒ $\frac{1}{5}$

λ ⇒ num events Per Unit time

$\lambda = \frac{1}{5}$, $\beta = 5$

Q1 Find the probability of debugging in 4 to 5 minutes

$P(4 < T < 5)$

$P(T < 5) - P(T < 4)$

exp.cdf(5) — exp.cdf(4)

≈ 8.1 %

Q2 Find the probability of needing more than 6 minutes to debug

$P(T > 6) = 1 - P(T \leq 6)$

$\boxed{1 - \text{e.cdf}(6)}$

≈ 30.11 %

Q3. Given that you have already spent 3 minutes, what is the probability of needing more than 9 minutes

$$P(T>9 \mid T>3) \Rightarrow$$

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

$$P(T>9) \cap P(T>3) \Rightarrow P(T>9)$$

since everything >9 will be >3 as well

$$\frac{P(T>9) \cap P(T>3)}{P(T>3)} \Rightarrow \frac{P(T>9)}{P(T>3)}$$

$$\frac{P(T>9)}{P(T>3)} \Delta \frac{1-e.cdf(9)}{1-e.cdf(3)}$$

$$\downarrow$$

$$P(T>6)$$

The first 3 seconds had no impact on next 6 seconds

Memoryless property of Exponential Distribution

$\Rightarrow$ Probability of Needing more time in future is Same Regardless of How much time has already passed.

Exp-Dist Treats every moment as new moment

Quiz

Suppose you have a system that fails, on average, every 50 hours. What is the probability that the system will fail within the first 20 hours?

$50 \rightarrow ?$
β

$P(T \le 20)$
$\downarrow$
e.cdf(20, β)

Box-Cox Transform

log-transformation
Right Skewed $\longrightarrow$ Normal

Q Do we have a transformation that can convert any Non-Normal Dist to Normal Dist

$$Y(\lambda) = \boxed{\frac{Y^{\lambda} - 1}{\lambda}} \quad \text{if } \lambda \neq 0$$

$$\boxed{\ln(Y)} \quad \text{if } \lambda = 0$$

* In Box-cox transformation, we have to find (λ)

$\lambda \Rightarrow$ Transformation Parameter

$Y \Rightarrow$ Original Data/Distribution

$Y(\lambda) \Rightarrow$ Transformed Distribution

Q How do we find λ

ML 1) Maximum Likelihood Estimation

ML 2) Grid Search

Stats 3) QQ plot

ML 4) Cross-Validation

Scipy → Box-Cox
(MLE)

Geometric Distribution

Questions

```
Imagine you're in a job search, and you're giving interviews until you land your first job.
```

Q. What are the possible outcomes in this situation?

S (p)   F (1-p)

$P_S = 0.1$
$P_F \Rightarrow 0.9$

1 Interview

$S \Rightarrow P(S) \Rightarrow 0.1 \Rightarrow F^0 S^1$

2 Interviews

$P(F\,S) \Rightarrow 0.9 * 0.1 \Rightarrow F^1 S^1$

3 Interviews $\Rightarrow$ $0.9^2 \times 0.9 * 0.9 \Rightarrow F^2 S^1$

$P(F\,F\,S)$

$\Rightarrow$ Getting First after N failure

$F^n \quad S^1$

$\Rightarrow$ Getting first success in K attempts

$F^{k-1} * S$

Getting success in K attempt $\Rightarrow P_S$

$$\boxed{P(X = K) = (1 - P)^{K-1} * P}$$

Expected Value in Geometric Distribution

```
You are flipping a fair coin repeatedly until you get heads for the first time.
You're interested in finding out,
on average, how many times you need to flip the coin before you get that first heads.
```

$$\boxed{E(X) = \mu \Rightarrow \frac{1}{P_S}}$$

$\Rightarrow \frac{1}{2} \Rightarrow 0.5$

$\Rightarrow \frac{1}{1/2} \Rightarrow 2$ (on an avg you will need 2 trial to get First Succ)

Quiz

You are flipping a biased coin with a 30% chance of getting heads until you succeed.
What is the probability of getting heads on the 2nd flip?

$P = 0.3$

$k = 2$

g.pmf(k, p)

Quiz

In a factory that produces light bulbs, there's a 5% chance that any given bulb is defective.
What is the probability of needing 1 or 2 bulbs to find the first defective bulb?

$P = 0.05$

$k = 1$ or 2

g.pmf(1, p) + g.pmf(2, p)

g.cdf(2, p)

Confidence Interval with Bootstrap

CLT
↓
Sampling Distribution
↓
means of Multiple Samples
↓
CI

When to Use?

S.D.S. ⇒ [35, 36, 33, 37, 34]

⇒ We use Bootstrap only when we have very limited Data

Step 1: Bootstrap Sampling

⇒ where sample for given Data with Replacement

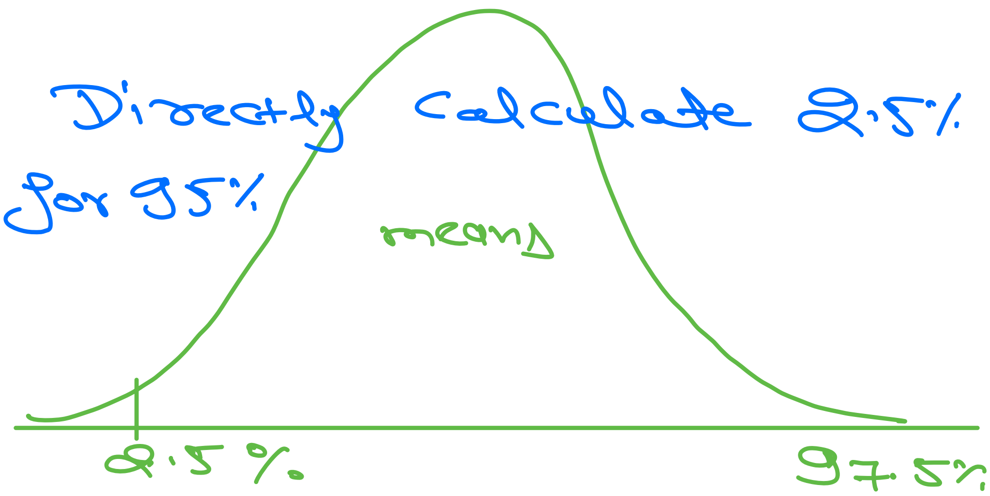
[35, 36, 33, 37, 34] → S_1 ⇒ [35, 36, 35, 35]

→ S_2 ⇒ [33, 34, 35, 35]

. . . S_n ⇒ . . . .

Step 2: Directly Calculate 2.5% and 97.5 for 95%

S.E X



⇒