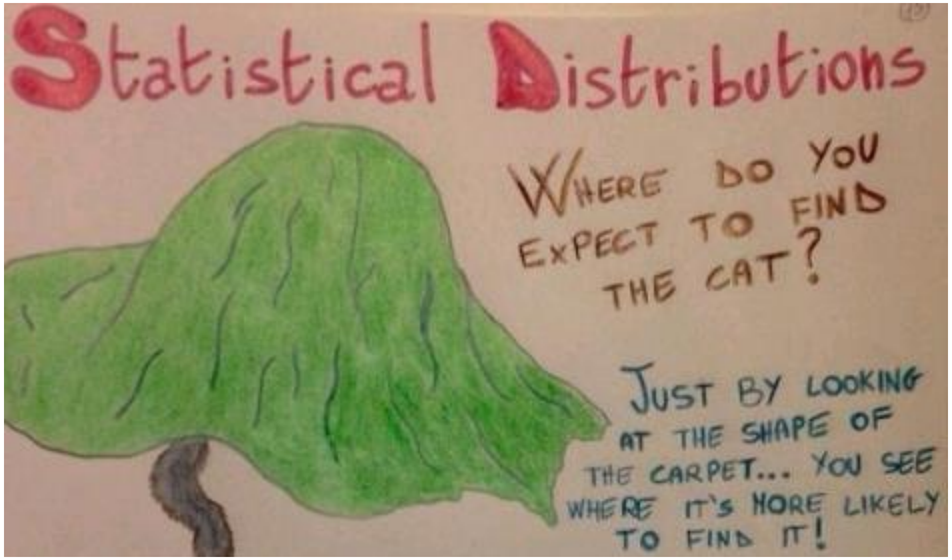
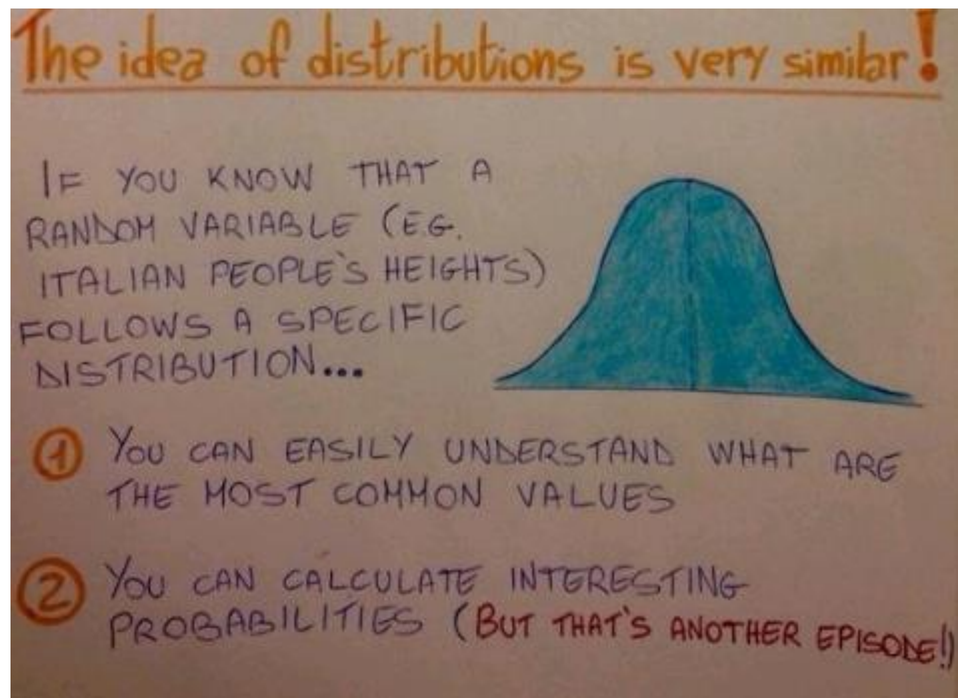
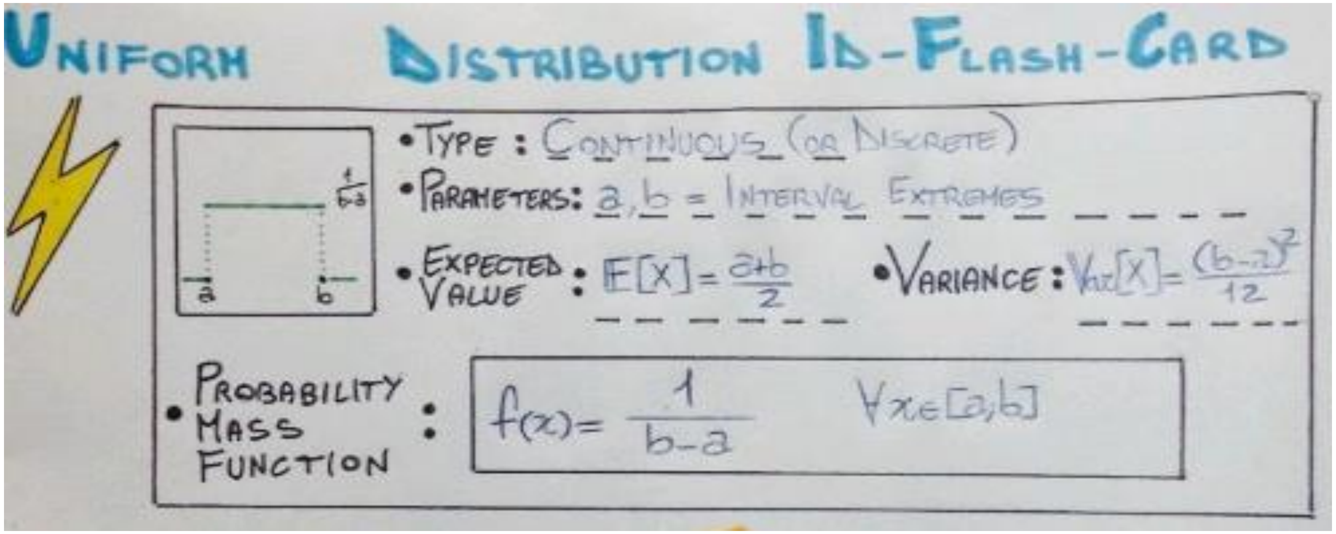
# Introduction to Distributions

# Statistical Distribution



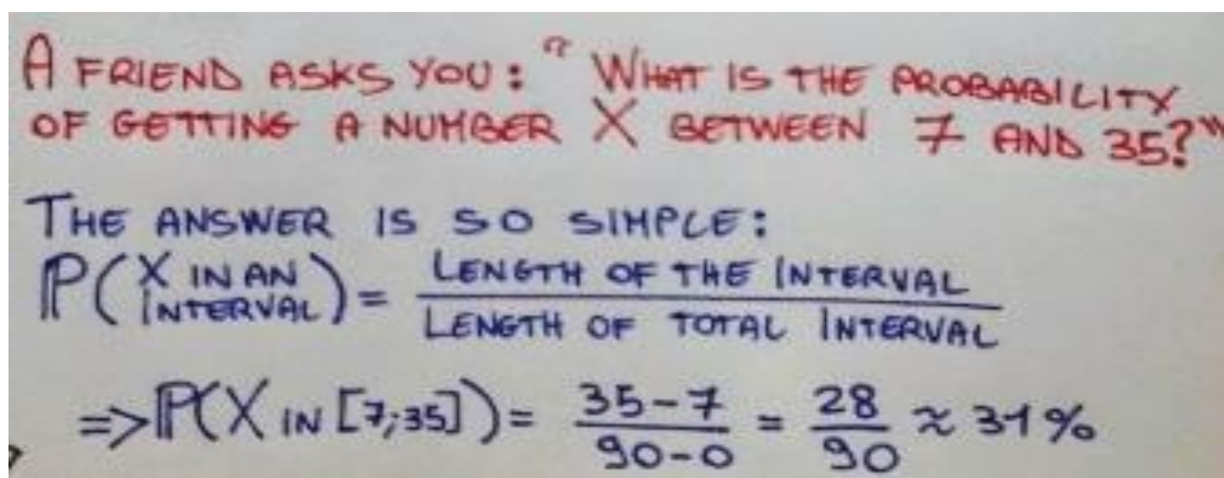



# Uniform Distribution

- The uniform distribution is used in situations where each outcome of a random experiment has same probability.
- It can be discrete or continues and its defined over an interval [a , b] corresponding to the set of possible outcomes.
- 

# Practical Example

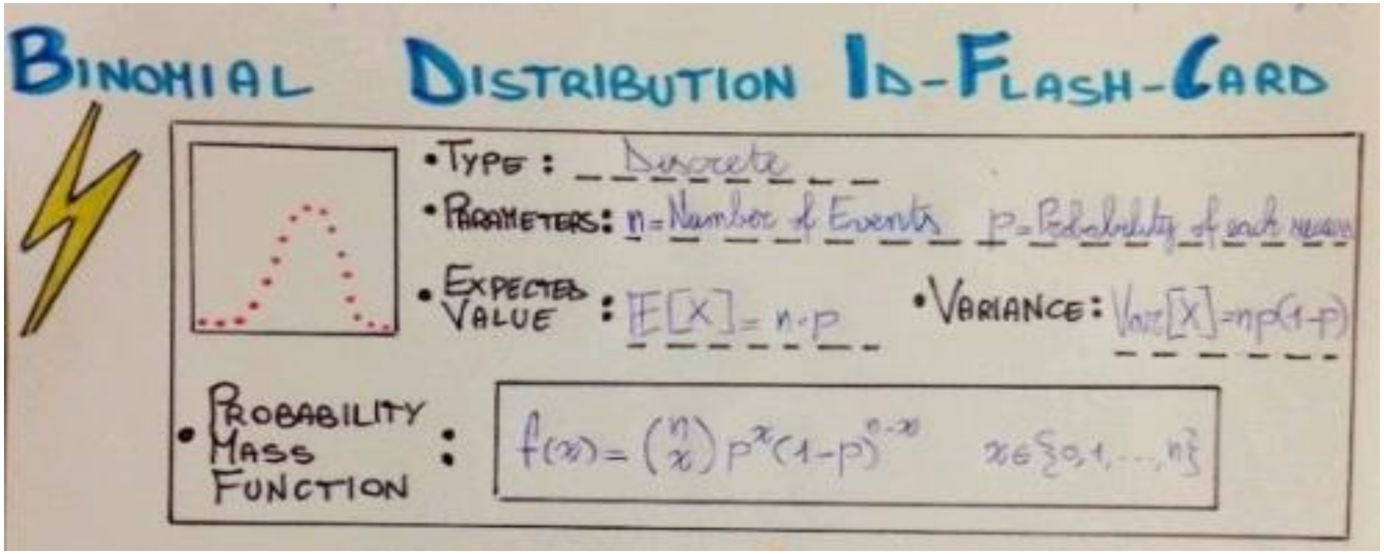
- Suppose your are playing A “Special” BINGO:
- The Bag contains the continuous set of numbers between 0 to 90. (Eg:1.2,2.4……………………….And so on).

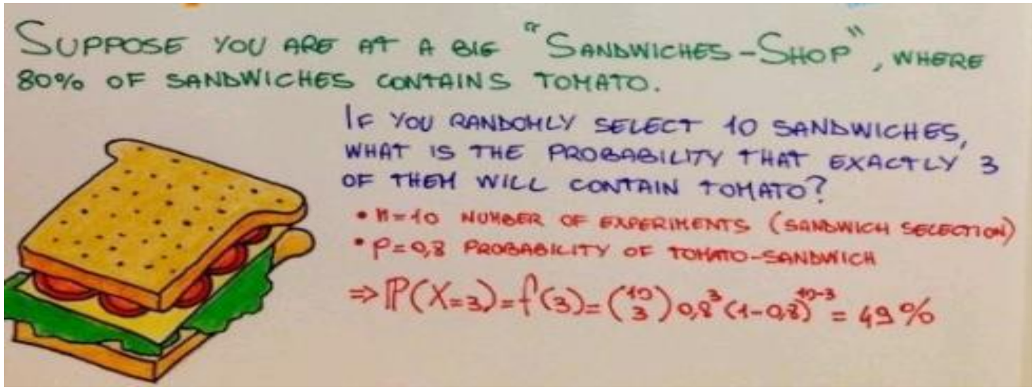
# Binomial Distribution

- Suppose the probability of a single trial being a success is p.
- Then the probability of observing exactly k successes in n independent trials is given by.

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$



# Practical Example




What is the probability that 3 of 8 randomly selected students will refuse to administer the worst shock, i.e. 5 of 8 will?

$$\begin{aligned}\binom{8}{3}(0.35)^3(1-0.35)^{8-3} &= \frac{8!}{3!(8-3)!}(0.35)^3(1-0.35)^{8-3} \\ &= \frac{8!}{3!5!}(0.35)^3(0.65)^5\end{aligned}$$

Dealing with the factorial part:

$$\frac{8!}{3!5!} = \frac{8\times7\times6\times5\times4\times3\times2\times1}{(3\times2\times1)(5\times4\times3\times2\times1)} = \frac{8\times7\times6}{3\times2\times1} = 56$$

Using $(0.35)^3(0.65)^5 \approx 0.005$, the final probability is about $56 * 0.005 = 0.28$.

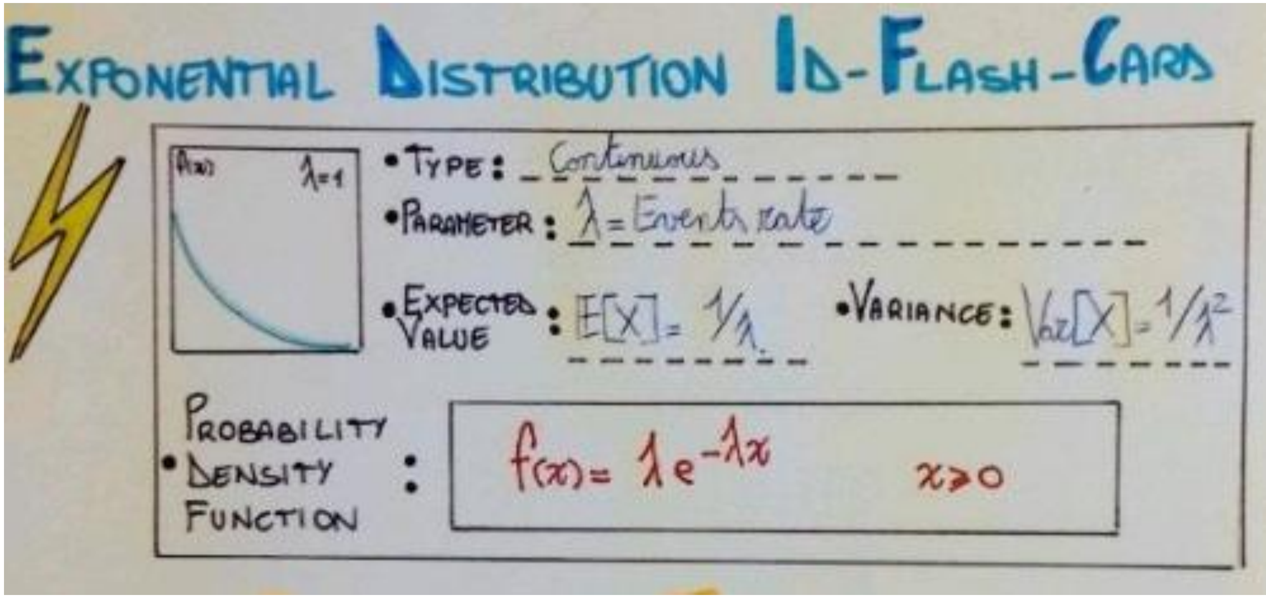
# Exponential Distribution

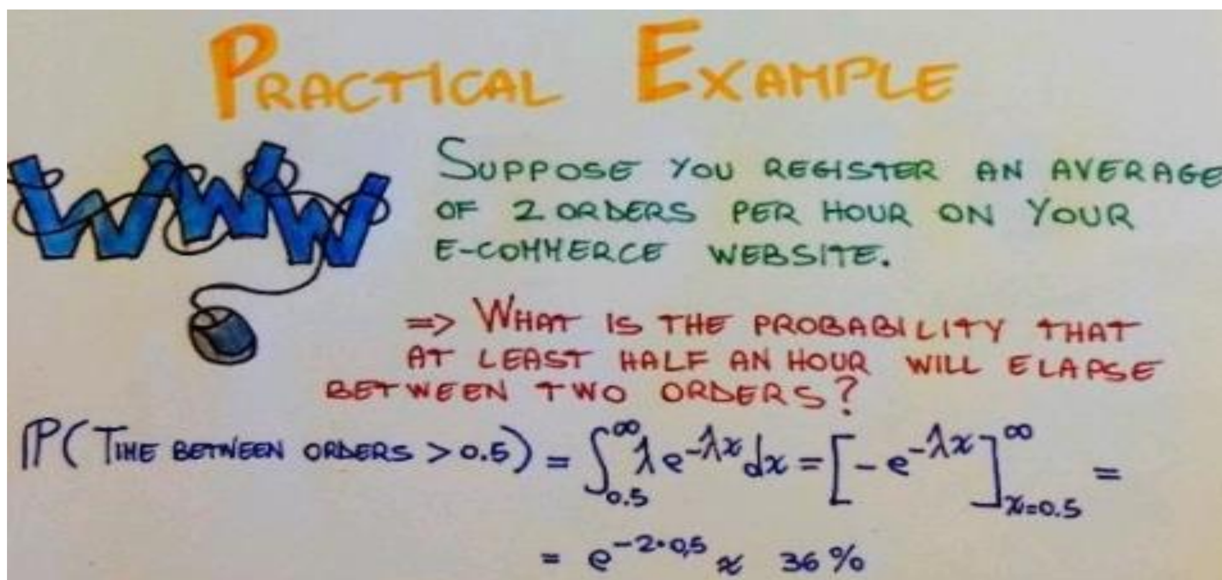
- The Exponential distribution describes the probability of the time between two events knowing that these events occurs independently with constant average frequency .

- Probability that a customer will visit in ndays

$$PDF = \lambda e^{-n\lambda}, n \geq 0$$

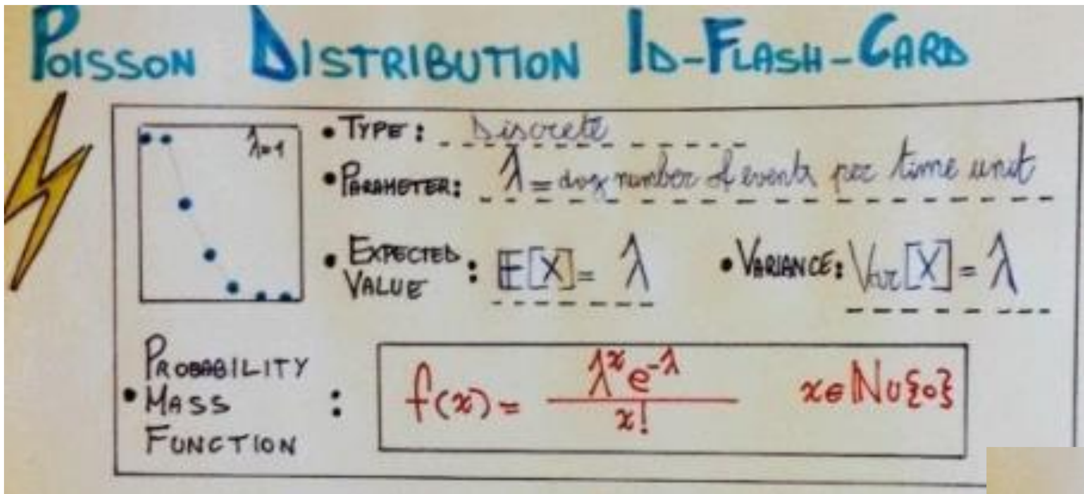
# Cont'd…

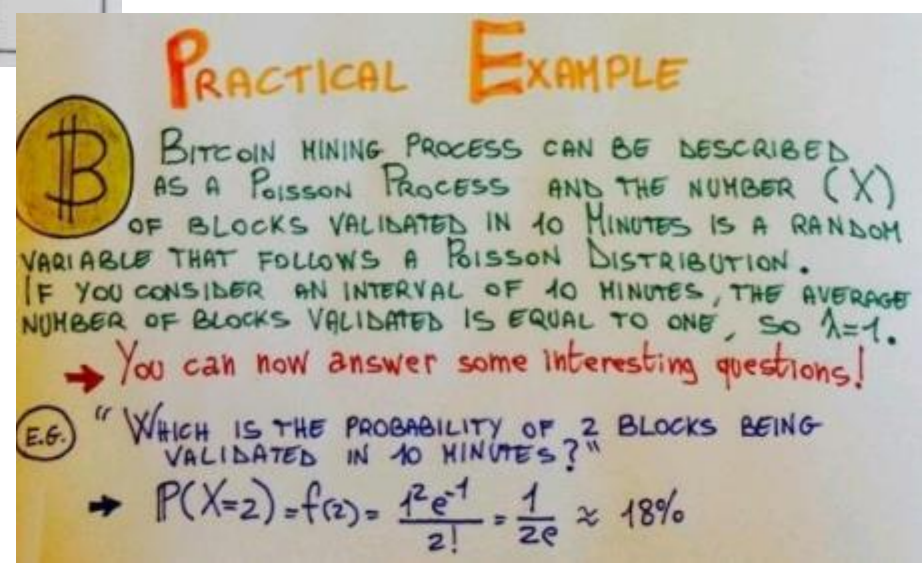

EXPONENTIAL DISTRIBUTION ID-FLASH-CARDS
f(x)
$\lambda=1$
• TYPE: Continuous
• PARAMETER: $\lambda$ = Events rate
• EXPECTED VALUE: $E[X] = 1/\lambda$
• VARIANCE: $Var[X] = 1/\lambda^2$
• PROBABILITY DENSITY FUNCTION: $f(x) = \lambda e^{-\lambda x}$ $x \geq 0$


PRACTICAL EXAMPLE
SUPPOSE YOU REGISTER AN AVERAGE OF 2 ORDERS PER HOUR ON YOUR E-COMMERCE WEBSITE.
=> WHAT IS THE PROBABILITY THAT AT LEAST HALF AN HOUR WILL ELAPSE BETWEEN TWO ORDERS?
$\mathbb{P}(\text{TIME BETWEEN ORDERS} > 0.5) = \int_{0.5}^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_{x=0.5}^{\infty} =$
$= e^{-2 \cdot 0.5} \approx 36\%$

# Poission Distribution

- The Possion Distribution Describe the probability of a number of events occurring in given time interval.



# Geometric Distribution

- Number of independent and identical (i.i.d) Bernoulli trials needed to get ONE success, e.g., number of people I need to call for the first person to accept loan.
- If the probability of a success in one trial is p and the probability of a failure is 1 – p, then the probability of finding the first success in the nth trial is given by

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

# Probability Distribution(Discrete)

- <u>Geometric</u>: For estimating number of attempts before first success
- <u>Binomial</u>: For estimating number of successes in *n* attempts
- <u>Poisson</u>: For estimating n number of events in a given time period when on average we see m events

# Probability Distributions –Scenarios

- Identify the distribution and calculate expectation, variance and the required probabilities.

Q1. A man is bowling. The probability of him knocking all the pins over is 0.3. If he has 10 shots, what is the probability he will knock all the pins over less than 3 times?

Binomial Distribution

Q2. On average, 1 bus stops at a certain point every 15 minutes. What is the probability that no buses will turn up in a single minute interval?

Poission Distribution.

Q3. 20% of cereal packets contain a free toy. What is the probability you will need to open fewer than 4 cereal packets ?

$X \sim Geo(0.2)$; $p=0.2$, $q=1-0.2=0.8$, $r<4$ or $\leq 3$

$E(X) = 1/p = 5$

$Var(X) = q/p2 = 20$

$PX \leq r = 1 - qr$

$P(X \leq 3) = 0.488$

Q4.Suppose 14 students each have a .6 probability of passing statistics. What’s the probability that 3 or more will pass?

- **Binomial Distribution**
- p = 0.6, q=0.4, n=14, r=3
- P(X>=3) = 1−P(X<3)
- P(X>=3) = 1−[P(X=0) + P(X=1) + P(X=2)]
- $PX=r=nCrprqn-r$
- P(X>=3) = 1 −[ 0.0006]
- P(X>=3) = 0.9994
- 0.000608677

- Suppose births in a hospital occur randomly at an average rate of 1.8 births per hour. What is the probability of 5 births in a given 2 hour interval?

What is $\lambda$?

$$P(X=5)=\frac{e^{-3.6}3.6^5}{5!}\ or\ \frac{e^{-1.8*2}(1.8*2)^5}{5!}\ ?$$

# Poisson or Exponential?

- The *number* of events in a given time period — Poisson
- The *time* until the first event
- The *time* from now until the next occurrence of the event
- The *time interval* between two successive events

Exponential

A tech support center of a company receives 5 calls per hour on an average.

a) What is the probability that the center will receive 8 calls in the next hour?
b) What is the probability that more than 30 minutes will elapse between calls?
c) What is the probability that more than 30 minutes and less than 45 minutes will elapse between calls?

**a) Poisson Distribution**

$$P(X = 8) = \frac{e^{-5}5^{8}}{8!} = 0.065$$

**b) Exponential Distribution**

$$P(Time\ between\ calls > 30) = \int_{0.5}^{\infty} \lambda e^{-\lambda T} dT = -e^{-\lambda T}\Big]_{0.5}^{\infty} = e^{-5*0.5}$$

$$= 0.082$$

**c) Exponential Distribution**

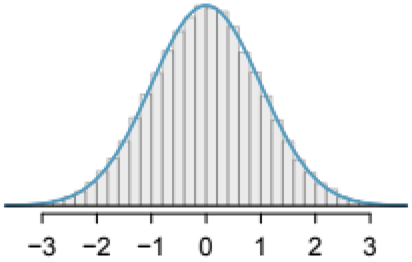
$$P(Time\ between\ calls > 30\ and < 45) = \int_{0.5}^{0.75} \lambda e^{-\lambda T} dT = -e^{-\lambda T}\Big]_{0.5}^{0.75}$$

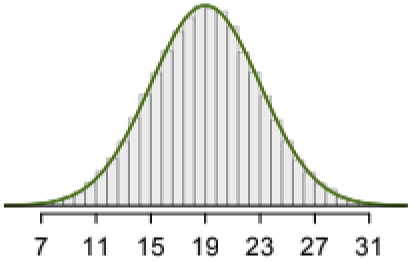
$$= -e^{-5*0.75} + e^{-5*0.5}$$

$$=0.058$$

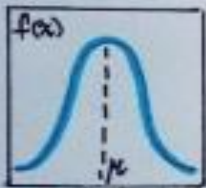
# Normal Distribution

- The normal distribution model always describes a symmetric, unimodal, bell-shaped curve.
- Specifically, the normal distribution model can be adjusted using two parameters: mean and standard deviation.



# GAUSSIAN DISTRIBUTION ID-FLASH-CARD




- TYPE: Continuous
- PARAMETERS: $\mu$ = Mean $\sigma^2$ = Variance
- EXPECTED VALUE: $\mathbb{E}[X] = \mu$
- VARIANCE: $\mathrm{Var}[X] = \sigma^2$
- PROBABILITY DENSITY FUNCTION:

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in \mathbb{R}}$$

# PRACTICAL EXAMPLE

A COMPANY PRODUCES SUGAR BAGS OF 500 g AND, IF YOU CHECK A SAMPLE OF BAGS, YOU'LL SEE THAT WEIGHTS ARE NORMALLY DISTRIBUTED WITH A MEAN OF 502 g AND A STANDARD DEVIATION OF 3 g.

=> WHAT IS THE PROBABILITY THAT A RANDOM PICKED BAG IS UNDER 495 g?

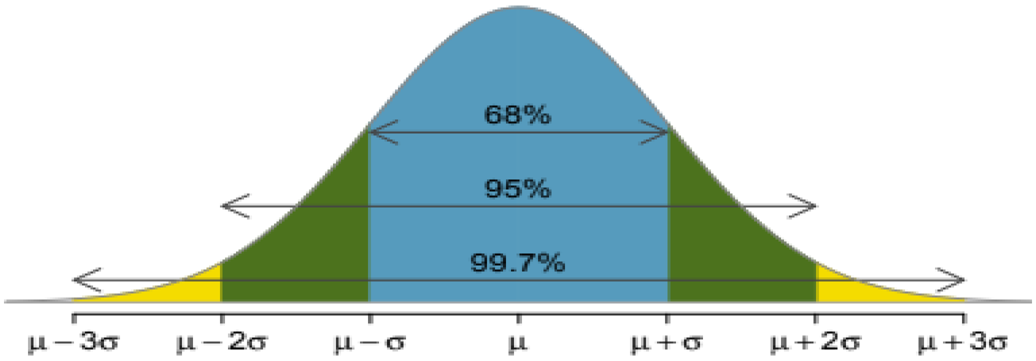
- STANDARDIZE $x=490 \rightarrow z = \frac{495-502}{3} = -2.33$
- USE STANDARD GAUSSIAN TABLE TO CALCULATE:
  $\mathbb{P}(Z < -2.33) = \Phi(-2.33) \simeq 1\%$

# Normal Distribution

**The Z-score**

The Z-score of an observation is the number of standard deviations it falls above or below the mean. We compute the Z-score for an observation $x$ that follows a distribution with mean $\mu$ and standard deviation $\sigma$ using

$$Z = \frac{x - \mu}{\sigma}$$

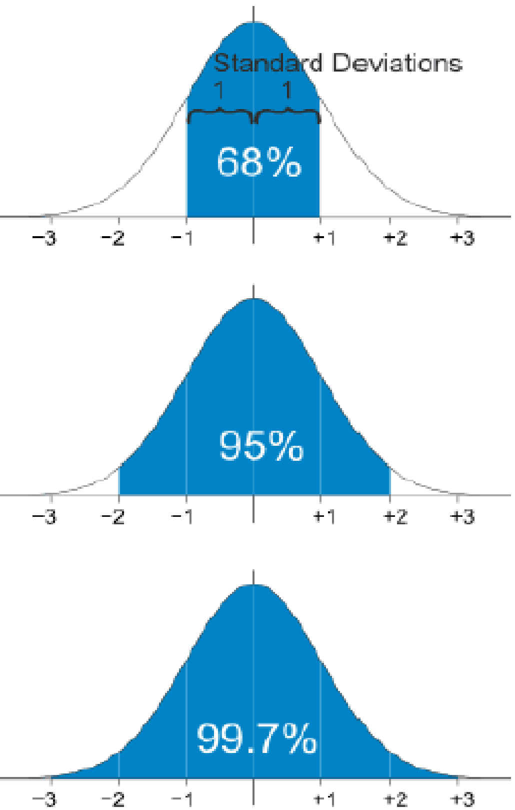


# 68-95-99.7 rule

- If the data is normally distributed then
- 68% of the data is within the ±one standard deviation (±1s) from the mean
- 95% of the data is within the ±two standard deviations (±2s) from the mean
- 99.7% of the data is within the ±three standard deviations (±3s) from the mean

# Measures of Spread (Dispersion)

You know the 68-95-99.7 rule.




A company produces a valve that is specified to weigh 1500g, but there are imperfections in the process. While the mean weight is 1500g, the standard deviation is 300g.

Mean ($\mu$) = 1500g and Standard deviation ($\sigma$) = 300

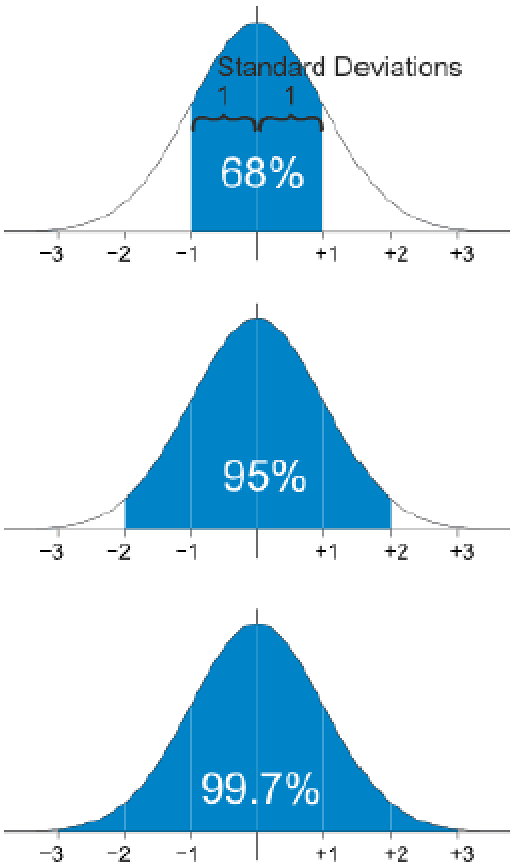
Q1. What is the range of weights within which 95% of the valves will fall?

Ans) 95% of the data will be between $\pm 2\sigma$ from the mean . So 1500 $\pm$ (2*300) = Between 900g and 2100g

Q2. Approximately 16% of the weights will be more than what value?

Ans) 32% of the data will be outside $\pm 1\sigma$ from the mean, because 68% of the data is between $\pm 1\sigma$ . Since it is symmetrical we have 16% of the weights on each side outside of $\pm 1\sigma$ . So 1500 + (1*300) = 1800 g. So 16% of the weights will be greater than 1800g

# You know the 68-95-99.7 rule.




A company produces a valve that is specified to weigh 1500g, but there are imperfections in the process. While the mean weight is 1500g, the standard deviation is 300g.

Mean ($\mu$) = 1500g and Standard deviation ($\sigma$) = 300

Q3. Approximately 0.15% of the weights will be less than what value?

Ans) 0.3% (100 – 99.7) of the data will be outside $\pm 3\sigma$ from the mean, because 99.7% of the data is between $\pm 3\sigma$ . Since it is symmetrical we have 0.15% of the weights on each side outside of $\pm 3\sigma$ . So 1500 - (3*300) = 600 g. So 0.15% of the weights will be less than 600g

# Summary

**Probability Distributions – Discrete**

| Distribution | Geometric | Binomial | Poisson | Exponential |
|---|---|---|---|---|
| Type | Discrete | Discrete | Discrete | Continuous |
| Representation | X~Geo(p) | X~B(n,p) | X~Po(λ) | X~Exp(λ) |
| Explanation | For estimating number of attempts before first success | For estimating number of successes in "n" attempts | For estimating "n" number of events in a given time period when on average we see "m" events | Time between events |
| Expected Value | $E(X) = \frac{1}{p}$ | $E(X) = np$ | $E(X) = \lambda$ | $E(X) = \frac{1}{\lambda}$ |
| Variance | $Var(X) = \frac{q}{p^2}$ | $Var(X) = npq$ | $Var(X) = \lambda$ | $Var(X) = \frac{1}{\lambda^2}$ |
| Probability Mass Function (PMF) P(X=x) | $P(X = r) = q^{r-1}p$ | $\mathrm{P}(\mathrm{X} = \mathrm{r}) = C_r^n p^r q^{n-r}$ | $\mathrm{P}(\mathrm{X} = \mathrm{r}) = \frac{e^{-\lambda}\lambda^r}{r!}$ | P(X=x) = $\lambda e^{-n\lambda}, n \geq 0$ |
| Cumulative Distribution Function (CDF). P(X<=x) | $P(X \leq r) = 1 - q^r$ | $\mathrm{P}(\mathrm{X} \leq \mathrm{r}) = \sum_{i=0}^{r} C_i^n p^i q^{n-i}$ | $\mathrm{P}(\mathrm{X} \leq \mathrm{r}) = e^{-\lambda} \sum_{i=0}^{r} \frac{\lambda^i}{i!}$ | $1 - e^{-n\lambda}, n \geq 0$ |
| P(X>x) = 1- P(X<=x) | $q^r$ | $1 - \sum_{i=0}^{r} C_i^n p^i q^{n-i}$ | $1 - e^{-\lambda} \sum_{i=0}^{r} \frac{\lambda^i}{i!}$ | $e^{-n\lambda}, n \geq 0$ |

- END