# Probability & Statistics for Machince Learning & Data Science

## Week 1: Introduction to Probability and Probability Distributions

The probability of occurance of an event within a sample space of choices is:

$$P(E) = \frac{\text{no. of Event}}{\text{no. of sample Space}}$$

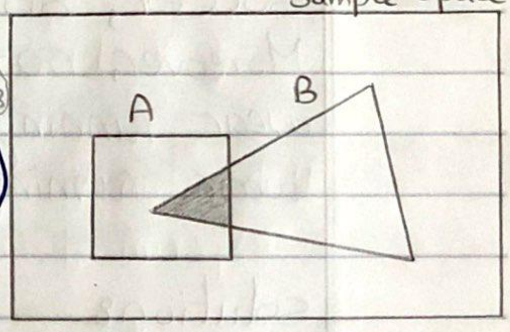
Complement Rule

The complement probability (probability of not occuring) is:

$$P(E') = 1 - P(E)$$

# Disjoint Events A & B → $P(A \cap B) = 0$
# Joint Events A & B → $0 < P(A \cap B) \leq 1$

Sample Space ($\Omega$)




### Sum Rule

(Added once with A & once with B)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

# OR → $\cup$ # AND → $\cap$

→ Independence:

An event is independent of the other, when its occurence doesn't affect of the occurence of the other.

# If events X & Y are independent, then:

$$P(X \cap Y) = P(X) \cdot P(Y)$$

# General Product Rule

$$P(A \cap B) = P(A) \cdot \underbrace{P(B \mid A)}_{\text{conditional Probability}}$$

# The probability of B given that A is happening; means if A happens or occurs what is the probability of B. This is:

conditional

$$P(B \mid A) = \frac{\text{no. of B exist in A occurence}}{\text{no. of A}}$$

# if B & A are independent $P(B \mid A) = P(B)$

$P(A \cap B) = P(A) \cdot P(B)$

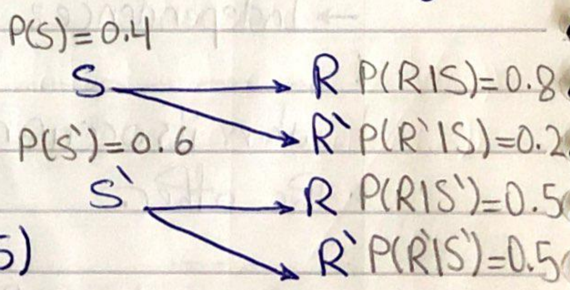
Example:

In a schod of 100 students, 40 of them play soccer. Among the students who play soccer, 80 % of them wear running shoes. Moreover, 50% of those of don't play soccer wear running shoes too. How many student wear running shoes?

solutions:

S → Play Soccer  R → wear running shoes

$$P(R) = P(S \cap R) + P(S' \cap R)$$
$$= P(S) \cdot P(R \mid S)$$
$$+ P(S') \cdot P(R \mid S')$$
$$= (0.4) \cdot (0.8) + (0.6) \cdot (0.5)$$
$$P(R) = 0.62$$




$$P(R) = \frac{\text{no. of R}}{\text{no. of students}} \rightarrow \text{no. of R} = \underline{62} \text{ Students}$$

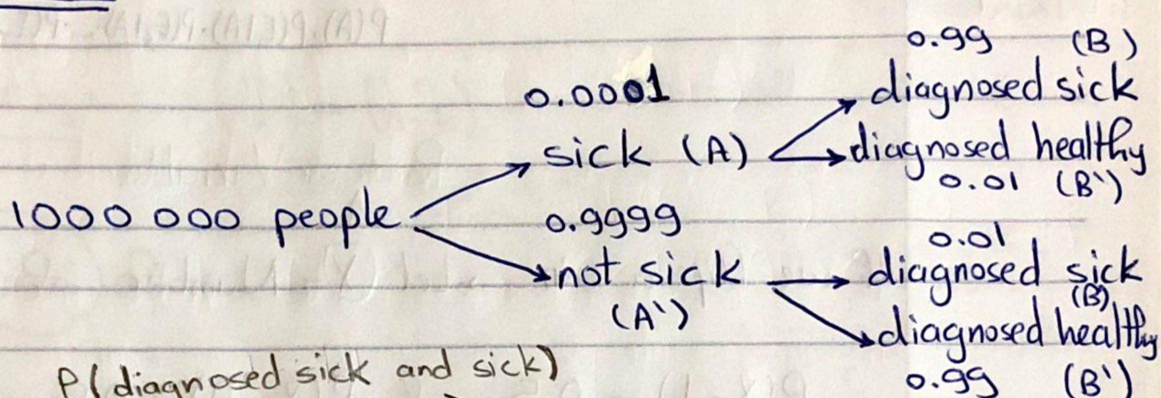
# From general product rule: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

$\boxed{\text{Bayes Theorem}}$: $P(A|B) = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(A') P(B|A')}$

Example:

In a population of 1000000 people, there is a rare disease that gets 1 in each 10000 people. Moreover, the diagnostic test is 99% efficient. Find the probability of being sick given ~~known~~ that diagnosed sick.

**solution:** A $\to$ Sick  B $\to$ dignosed sick




P(diagnosed sick and sick)

$$P(\text{sick} \mid \text{diagnosed sick}) = \frac{(0.0001)\cdot(0.99)}{(0.0001)\cdot(0.99) + (0.9999)\cdot(0.01)}$$

P(diagnosed sick)

$= 0.0098 = 0.98\%$

another solution:

no. of diagnosed sick $= (0.99 * \overset{\text{sick}}{100} = 99$ people)

$+ (0.01 * \underbrace{999900}_{\text{healthy}} = 9999$ people)

$= 10098$ people

no. of sick from those who are diagnosed sick = 99 people

$$P(\text{sick} \mid \text{diagnosed sick}) = \frac{\text{no. sick from diagnosed sick}}{\text{no. of diagnosed sick}}$$

$$= \frac{99}{10098} = 0.0098 = 0.98\%$$

⇒ In the previous example:

A → Prior: 100 out of 1000000 people are sick

E → Event: Diagnostic Test is 99% efficient

P(A|E) → **Posterior**: P(sick | diagnosed sick) = 0.0098

⇒ Naive Assumption:

It's assuming that events (E) being considered building the model are happening independently. This can ease math a lot, eventhough they are dependent in many cases.

Naive Assumption →

$$P(A|E_1 \& E_2 \& \_\_ \& E_n) = \frac{P(A) \cdot P(E_1|A) \cdot P(E_2|A) \cdot \_\_ \cdot P(E_n|A)}{P(A) \cdot P(E_1|A) \cdot P(E_2|A) \cdot \_\_ \cdot P(E_n|A) + P(A^c) \cdot P(E_1|A^c) \cdot P(E_2|A^c) \cdot \ldots \cdot P(E_n|A^c)}$$

Random Variable

If we flip a coin → Let $X$ = Number of heads $\begin{cases} X = 1 \\ X = 0 \end{cases}$

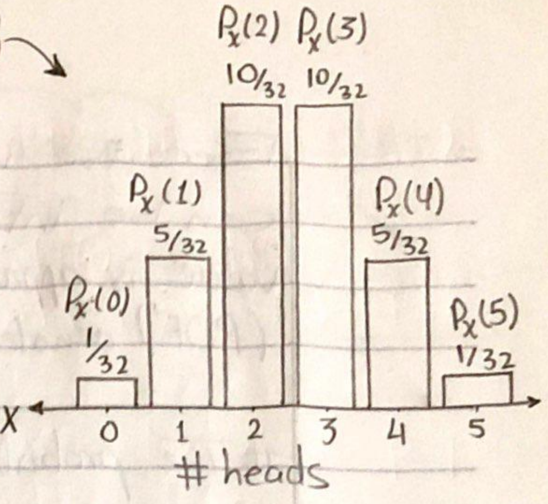
$P(X=1) = 0.5$ & $P(X=0) = 0.5$

# Random Variables allow you to model the whole experiment at once.

| Discret Random Variables | Continuous Random Variables |
|---|---|
| ~~((Finite Number of Values))~~ Not Precise | ~~((Infinite Number of Values))~~ |
| ((Can take only Countable Number of values)) | ((Take values on an interval)) |
| can be put in a list | |

⇒ Variables
- → Deterministic: take fixed outcomes
  $x = 2$, $f(x) = x^2$
- → Random: take uncertain outomes
  X = number of defective item in a shipment

Histogram




$X \rightarrow$ Number of heads in 5 coin tosses

⊛ Probability Mass Function (PMF):

$X = 0, 1, 2, 3, 4, 5$ & $P_X(x) = P(X = x)$

$P_X(x) \geq 0$ & $\sum_x P_X(x) = 1$

# There are 10 ways to have 2 heads in 5 coin tosses:

Number of ways you can order 5 coins

$$\frac{5!}{2!\,(5-2)!} = \underline{10} = C^5_2 = \binom{5}{2} \Rightarrow \text{Binomial Coefficient (combination)}$$

Number of heads

Number of not heads

Property: $\binom{n}{k} = \binom{n}{n-k}$ counts all combinations

All possible orders

Probability of seeing $x$ heads

Binomial: $P_X(x) = \binom{5}{x} p^x (1-p)^{5-x}$, $x = 0, 1, 2, 3, 4, 5$

Probability of seeing $5-x$ not heads

Event $X = x$: $x$ is heads in 5 tosses

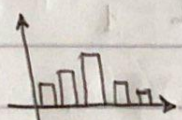
$\longrightarrow$ X follows a binomial distribution

$\longrightarrow$ $X \sim \text{Binomial}(5 \cdot p)$

Number of Flips

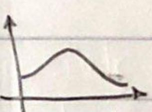
$P(H)$

Bernoulli:
- Success: Occurance of the preferable event ($X$)
  $p$
- Failure: Not occurance of the peferable event ($X'$)

Discrete 

# Sum of heights equals 1.

$$\sum_x P_X(x) = 1$$

Continuous 

# Area under curve equals 1.

$$\int_x P_X(x)\, dx = 1$$

Because in continuous random variables number of values can be infinite, and the probability of an exact value is approximately zero. The probability density function (PDF) denoted as $f_X(x)$ uses intervals of $x$ instead.

# The probability density function $f_X(a \le x \le b) = \int_a^b p_X(x)\,dx$

# The commulative distribution function $f_X(x \le a) = \int_0^a p_X(x)\,dx$

$0 \le \text{CDF} \le 1$ and denoted as $F_X(x)$.

→ It is a curve that starts from zero to 1 at the end; where zero and one are the heights.

→ $F_X(x)$ can never decrease, it only increases till 1.

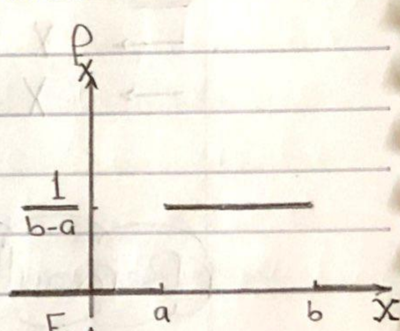
## Uniform Distribution

A continuous random variable can be modelled with it, if all possible values lie in an interval have the same frequency of occurance. Its parameters are:

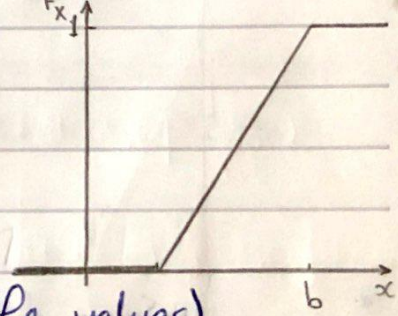
→ $a$: beginning of the interval

→ $b$: end of the interval

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x \notin (a,b) \end{cases}$$




$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & b \le x \end{cases}$$

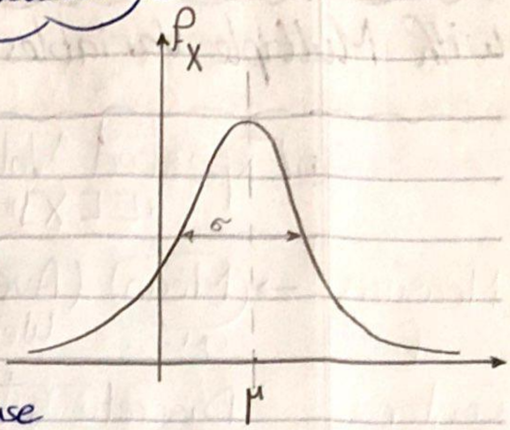



# mean → $\mu$ (average) $\cong$ (center of the values)

# standard deviation → $\sigma$ (measures the spread of values)

# Normal (Gaussian) Distribution

\# It is a distribution that is symmetric and takes the bell-shaped.




\# The closest function to this curve is $e^{-\frac{x^2}{2}}$. But because it doesn't fit well, we use this formula instead:

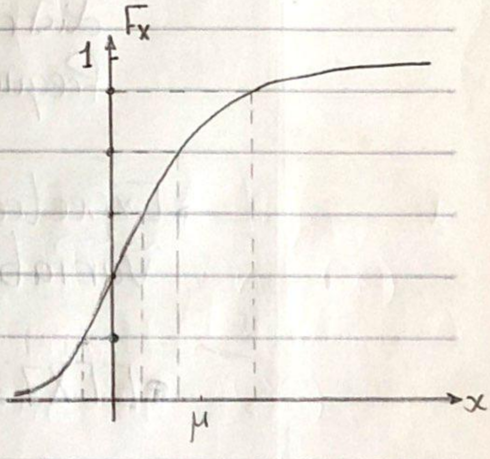
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \rightarrow X \sim \mathcal{N}(\mu, \sigma^2)$$

\# Standardization: The way of transforming any normal distribution to a standard one.

→ we calculate $Z = \frac{x-\mu}{\sigma}$, then the new $\mu = 0$ & $\sigma = 1$

$$f_X(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

\# Most of the natural phenomena can be represented using normal distribution.




→ To sample from distributions the y-axis of the CDF ($F_X$) gets divided into the number of samples, then these values gets translated into the sample values.

\# $\text{erf}(x) = \frac{1}{\sqrt{\pi/2}}\int_0^x e^{-t^2 \cdot 0.5}\,dt$

\# erf → error function

\# erfinv is from scipy.special

## Code

**Gaussian**

$$y = F(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$x = F^{-1}(y) = \left[\sigma\sqrt{2} \cdot \text{erf}^{-1}(2y-1)\right] + \mu$$

**Binomial**

$$y = F(x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$$

$0 \leq x \leq n$

$x = F^{-1}(y) =$ scipy.stats.binom.ppf(y, n, p)