312-3302 ARTIFICIAL INTELLIGENCE

Lecture 2

Mathematics

Review

Matrix

A Matrix is an array of numbers:

$$\begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}$$

A Matrix

(This one has 2 Rows and 3 Columns)

$A_{m \times n}$ หมายถึง matrix ขนาด m แถว n หลัก

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$\begin{array}{c c} & \begin{array}{cccc} 1 & 2 & \cdots & n \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array} & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \end{array}$$

Operation

Adding

To add two matrices: add the numbers in the matching positions:

3+4=7

$$\begin{bmatrix} 3 & 8 \\ 4 & 6 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 1 & -9 \end{bmatrix} = \begin{bmatrix} 7 & 8 \\ 5 & -3 \end{bmatrix}$$

These are the calculations:

3+4=7	8+0=8
4+1=5	6−9=−3

Operation

Subtracting

To subtract two matrices: subtract the numbers in the matching positions:

3−4 = −1

$$\begin{bmatrix} 3 & 8 \\ 4 & 6 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 1 & -9 \end{bmatrix} = \begin{bmatrix} -1 & 8 \\ 3 & 15 \end{bmatrix}$$

These are the calculations:

$3-4=-1$	$8-0=8$
$4-1=3$	$6-(-9)=15$

*Note: subtracting is actually defined as the **addition** of a negative matrix: $A + (-B)$*

Matrix

Operation

Multiply by a Constant

We can multiply a matrix by a **constant** *(the value 2 in this case)*:

$$2 \times \begin{bmatrix} 4 & 0 \\ 1 & -9 \end{bmatrix} \xrightarrow{2x4=8} \begin{bmatrix} 8 & 0 \\ 2 & -18 \end{bmatrix}$$

These are the calculations:

$2\times4=8$	$2\times0=0$
$2\times1=2$	$2\times-9=-18$

We call the constant a **scalar**, so officially this is called "scalar multiplication".

Operation

Matrix

Multiplying a Matrix by Another Matrix

But to multiply a matrix **by another matrix** we need to do the "dot product" of rows and columns ... what does that mean? Let us see with an example:

To work out the answer for the **1st row** and **1st column**:

"Dot Product"

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix} = \begin{bmatrix} 58 & \\ & \end{bmatrix}$$

To multiply an **m×n** matrix by an **n×p** matrix, the **n**s must be the same, and the result is an **m×p** matrix.

$$m \times n \;\times\; n \times p \;\rightarrow\; m \times p$$

Why Do It This Way?

This may seem an odd and complicated way of multiplying, but it is necessary!

I can give you a real-life example to illustrate why we multiply matrices in this way.

Example: The local shop sells 3 types of pies.

- Apple pies cost **$3** each
- Cherry pies cost **$4** each
- Blueberry pies cost **$2** each

And this is how many they sold in 4 days:

	Mon	Tue	Wed	Thu
Apple	13	9	7	15
Cherry	8	7	4	6
Blueberry	6	4	0	3

Now think about this ... the **value of sales** for Monday is calculated this way:

Apple pie value + Cherry pie value + Blueberry pie value

$3×13 + $4×8 + $2×6 = $83

So it is, in fact, the "dot product" of prices and how many were sold:

($3, $4, $2) • (13, 8, 6) = $3×13 + $4×8 + $2×6
= $83

We **match** the price to how many sold, **multiply** each, then **sum** the result.

Matrix

Operation

$$AB \neq BA$$

Example:

See how changing the order affects this multiplication:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1\times2+2\times1 & 1\times0+2\times2 \\ 3\times2+4\times1 & 3\times0+4\times2 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 10 & 8 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 2\times1+0\times3 & 2\times2+0\times4 \\ 1\times1+2\times3 & 1\times2+2\times4 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 7 & 10 \end{bmatrix}$$

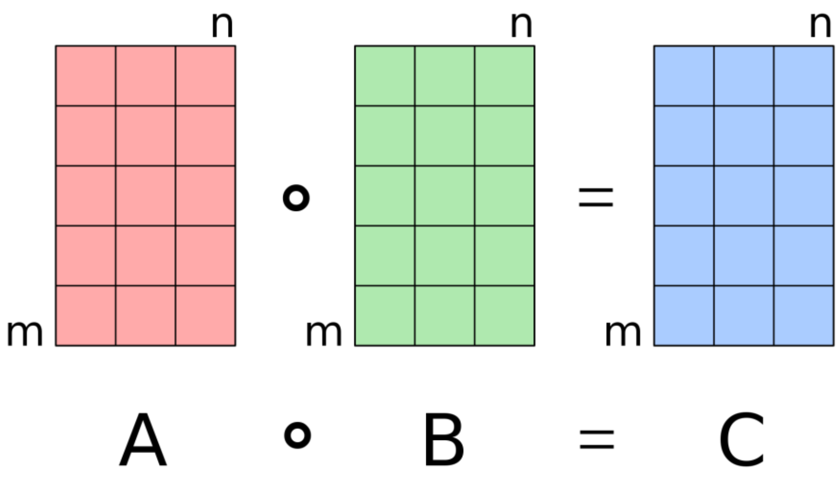
The answers are different!

Matrix

Operation

การคูณแบบ Hadamard product (element-wise product)

$$\begin{bmatrix} 2 & 3 & 1 \\ 0 & 8 & -2 \end{bmatrix} \circ \begin{bmatrix} 3 & 1 & 4 \\ 7 & 9 & 5 \end{bmatrix} = \begin{bmatrix} 2\times 3 & 3\times 1 & 1\times 4 \\ 0\times 7 & 8\times 9 & -2\times 5 \end{bmatrix} = \begin{bmatrix} 6 & 3 & 4 \\ 0 & 72 & -10 \end{bmatrix}$$



Matrix

Operation

Transposing

To "transpose" a matrix, swap the rows and columns.

We put a "T" in the top right-hand corner to mean transpose:

$$\begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}^{T} = \begin{bmatrix} 6 & 1 \\ 4 & -9 \\ 24 & 8 \end{bmatrix}$$

Matrix

Identity Matrix

An **Identity Matrix** has **1**s on the main diagonal and **0**s everywhere else:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A 3×3 Identity Matrix

- It is square (same number of rows as columns)
- It can be large or small (2×2, 100×100, ... whatever)
- Its symbol is the capital letter **I**

It is the matrix equivalent of the number "1", when we multiply with it the original is unchanged:

$$A \times I = A$$

$$I \times A = A$$

Matrix

Diagonal Matrix

A diagonal matrix has zero anywhere not on the main diagonal:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A diagonal matrix

Scalar Matrix

A scalar matrix has all main diagonal entries the same, with zero everywhere else:

$$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

A scalar matrix

Matrix

Triangular Matrix

Lower triangular is when all entries above the main diagonal are zero:

$$\begin{bmatrix} 5 & 0 & 0 \\ 2 & 1 & 0 \\ 7 & 6 & -3 \end{bmatrix}$$

A lower triangular matrix

Upper triangular is when all entries below the main diagonal are zero:

$$\begin{bmatrix} 2 & -2 & 7 \\ 0 & 4 & 11 \\ 0 & 0 & 5 \end{bmatrix}$$

An upper triangular matrix

Zero Matrix (Null Matrix)

Zeros just everywhere:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Zero matrix

Symmetric

In a Symmetric matrix matching entries either side of the main diagonal are **equal**, like this:

$$\begin{bmatrix} 3 & 2 & 11 & 5 \\ 2 & 9 & -1 & 6 \\ 11 & -1 & 0 & 7 \\ 5 & 6 & 7 & 9 \end{bmatrix}$$

Symmetric matrix

It must be square, and is equal to its own transpose

$$A = A^T$$

What is the Inverse of a Matrix?

Just like a **number** has a reciprocal ...

Reciprocal of a Number (note: $\frac{1}{8}$ can also be written 8^{-1})

... a **matrix** has an **inverse** :

Inverse of a Matrix

We write A^{-1} instead of $\frac{1}{A}$ because we don't divide by a matrix!

The inverse of A is A^{-1} only when:

$$AA^{-1} = A^{-1}A = \mathbf{I}$$

Sometimes there is no inverse at all.

What is the Inverse of a Matrix?

Just like a **number** has a reciprocal ...

Reciprocal of a Number (note: $\frac{1}{8}$ can also be written 8^{-1})

... a **matrix** has an **inverse** :

Inverse of a Matrix

We write A^{-1} instead of $\frac{1}{A}$ because we don't divide by a matrix!

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}^{-1} = \frac{1}{3\times8-4\times6}\begin{bmatrix} 8 & -4 \\ -6 & 3 \end{bmatrix}$$

$$= \frac{1}{24-24}\begin{bmatrix} 8 & -4 \\ -6 & 3 \end{bmatrix}$$

24−24? That equals 0, and **1/0 is undefined**.
We cannot go any further! This matrix has no Inverse.

Inverse Matrix

$$\begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}^{-1} = \frac{1}{3\times8-4\times6} \begin{bmatrix} 8 & -4 \\ -6 & 3 \end{bmatrix}$$

$$= \frac{1}{24-24} \begin{bmatrix} 8 & -4 \\ -6 & 3 \end{bmatrix}$$

24−24? That equals 0, and **1/0 is undefined**.
We cannot go any further! This matrix has no Inverse.

Such a matrix is called "Singular",
which only happens when the determinant is zero.

A Real Life Example: Bus and Train

A group took a trip on a **bus**, at $3 per child and $3.20 per adult for a total of $118.40.

They took the **train** back at $3.50 per child and $3.60 per adult for a total of $135.20.

How many children, and how many adults?

First, let us set up the matrices (be careful to get the rows and columns correct!):

$$\begin{array}{cc} \text{Child} & \text{Adult} \\ \end{array} \quad \begin{array}{cc} \text{Bus} & \text{Train} \\ \end{array} \quad \begin{array}{cc} \text{Bus} & \text{Train} \\ \end{array}$$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & 3.5 \\ 3.2 & 3.6 \end{bmatrix} = \begin{bmatrix} 118.4 & 135.2 \end{bmatrix}$$

$$XA = B$$

So to solve it we need the inverse of "A":

$$\begin{bmatrix} 3 & 3.5 \\ 3.2 & 3.6 \end{bmatrix}^{-1} = \frac{1}{3\times3.6-3.5\times3.2}\begin{bmatrix} 3.6 & -3.5 \\ -3.2 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} -9 & 8.75 \\ 8 & -7.5 \end{bmatrix}$$

Now we have the inverse we can solve using:

$$X = BA^{-1}$$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} 118.4 & 135.2 \end{bmatrix}\begin{bmatrix} -9 & 8.75 \\ 8 & -7.5 \end{bmatrix}$$

$$= \begin{bmatrix} 118.4\times-9 + 135.2\times8 & 118.4\times8.75 + 135.2\times-7.5 \end{bmatrix}$$

$$= \begin{bmatrix} 16 & 22 \end{bmatrix}$$

There were 16 children and 22 adults!

Probability

Probability

How **likely** something is to happen

Many events can't be predicted with total certainty. The best we can say is how likely they are to happen, using the idea of probability.

Tossing a Coin

When a coin is tossed, there are two possible outcomes:

Heads (H) or
Tails (T)

Also:

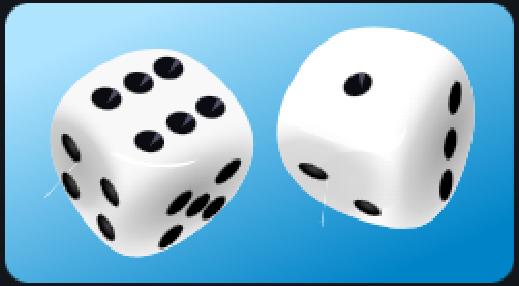
- the probability of the coin landing **H** is ½
- the probability of the coin landing **T** is ½

Probability

How **likely** something is to happen

Many events can't be predicted with total certainty. The best we can say is how likely they are to happen, using the idea of probability.

Throwing Dice



When a single die is thrown, there are six possible outcomes: **1, 2, 3, 4, 5, 6.**

The probability of any one of them is $\frac{1}{6}$

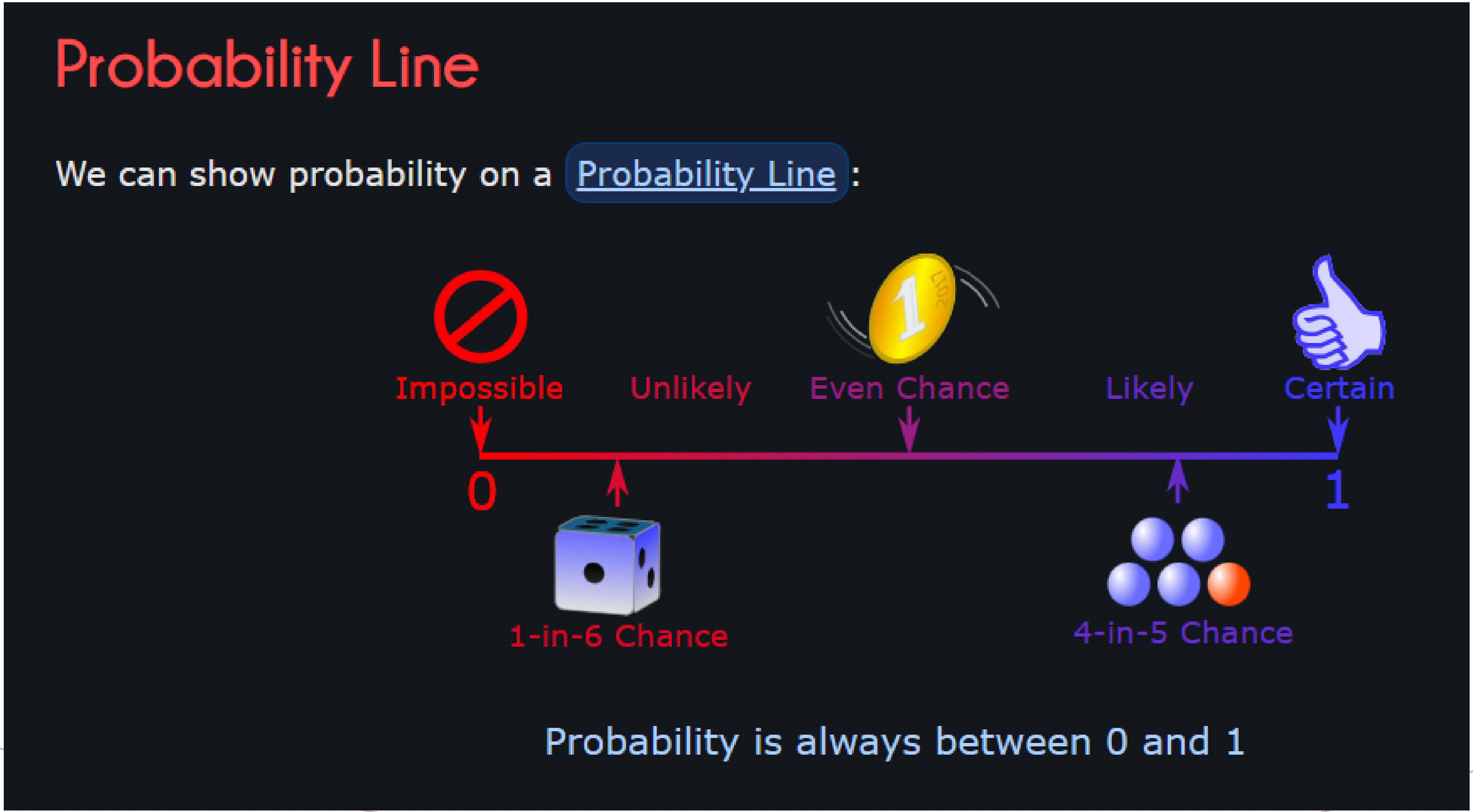
Probability

How **likely** something is to happen

Many events can't be predicted with total certainty. The best we can say is how likely they are to happen, using the idea of probability.

$$\text{Probability of an event happening} = \frac{\text{Number of ways it can happen}}{\text{Total number of outcomes}}$$

Probability



Probability

Experiment: a repeatable procedure with a set of possible results.

Example: Throwing dice

We can throw the dice again and again, so it is repeatable.

The set of possible results from any single throw is {1, 2, 3, 4, 5, 6}

Outcome: A possible result.

Example: "6" is one of the outcomes of a throw of a die.

Probability

Trial: A single performance of an experiment.

Example: I conducted a coin toss experiment. After 4 trials I got these results:

Outcome	Trial	Trial	Trial	Trial
Head	✓	✓		✓
Tail			✓	

Three trials had the outcome "Head", and one trial had the outcome "Tail"

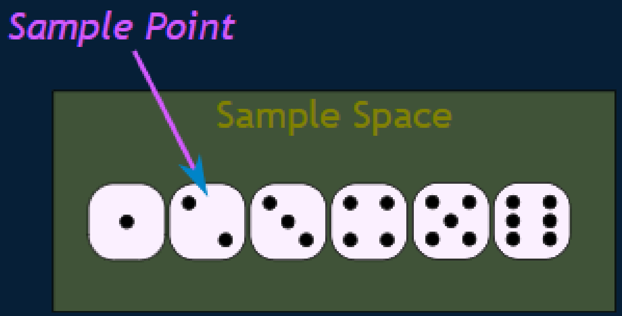
Probability

Sample Space: all the possible outcomes of an experiment.

Sample Point: just one of the possible outcomes

Example: Throwing dice

There are 6 different sample points in that sample space.



Probability

Event: one **or more** outcomes of an experiment

Example Events:

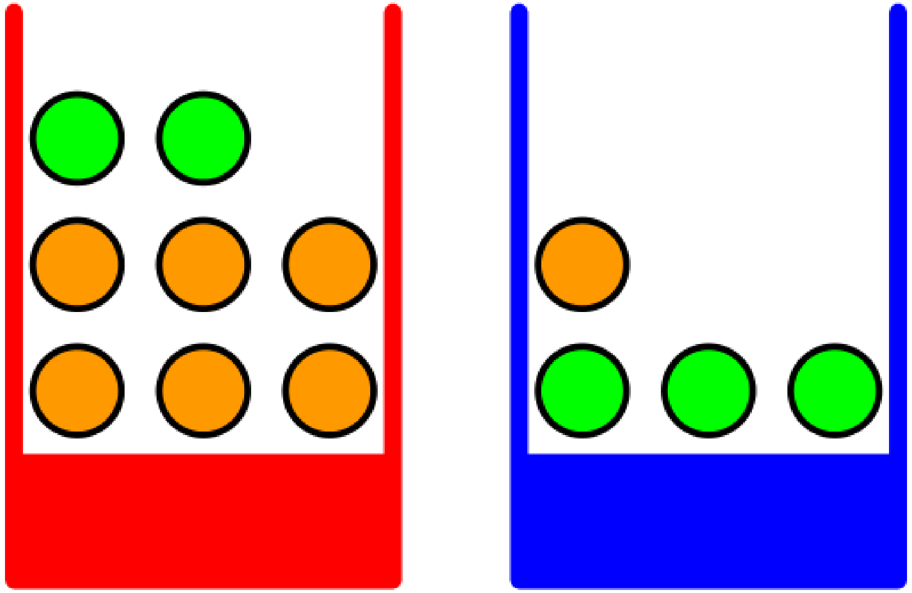
An event can be just one outcome:

- Getting a Tail when tossing a coin
- Rolling a "5"

An event can include more than one outcome:

- Choosing a "King" from a deck of cards (any of the 4 Kings)
- Rolling an "even number" (2, 4 or 6)

Probability

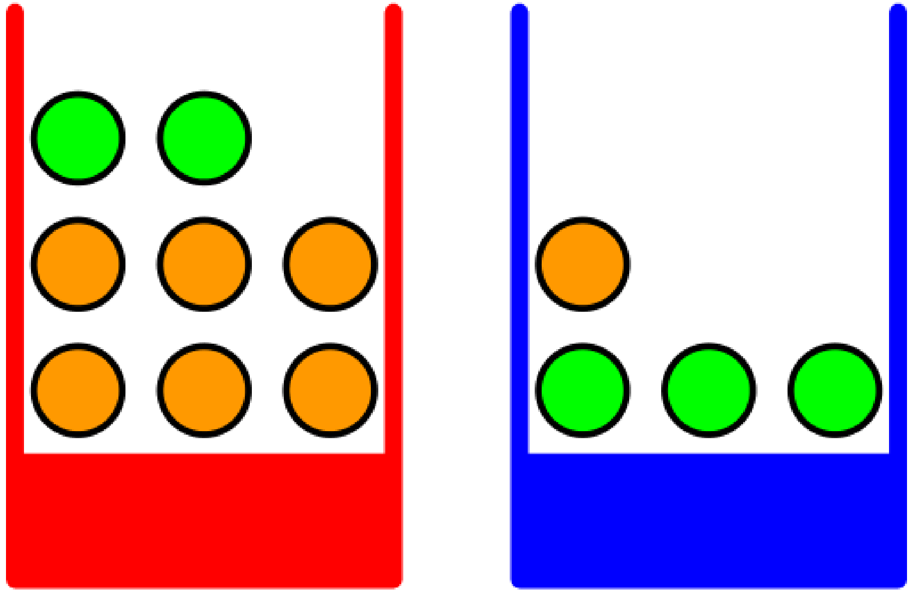


Box of Orange & Apple

Suppose we randomly pick one of the boxes and from that box we randomly select an item of fruit and having observed which sort of fruit it is we replace it in the box from which it came.

We could imagine repeating this process many times. Let us suppose that in so doing we pick the red box 40% of the time and we pick the blue box 60% of the time.

Probability



Box of Orange & Apple

In this example, the identity of the box that will be chosen is a **random variable**, which we shall denote by **B**

Two Possible values

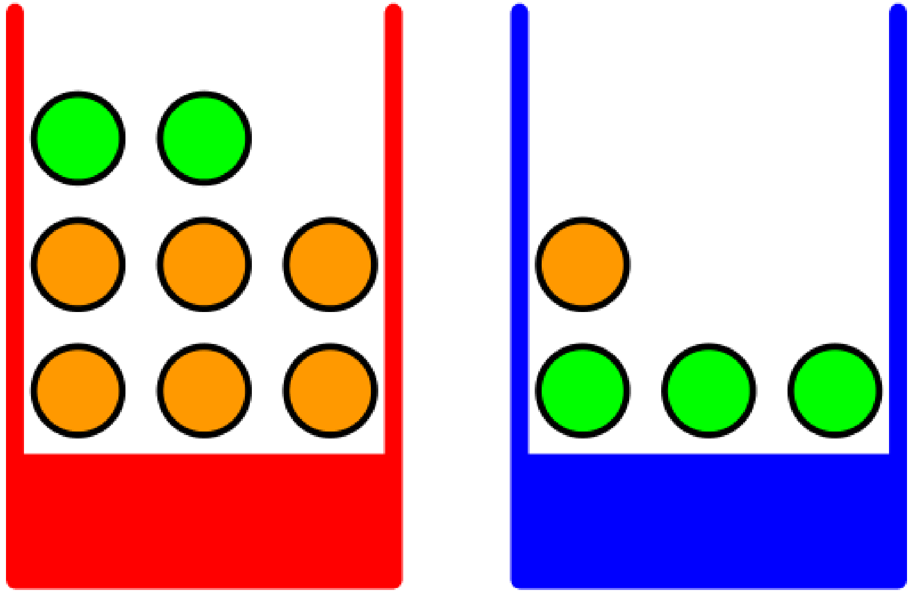
r (red box) and b (blue box)

Similarly, the identity of the fruit is also a **random variable** and will be denoted by **F**

Two Possible values

a (apple) and o (orange)

Probability



Box of Orange & Apple

ความน่าจะเป็นในการเลือกกล่องแต่ละสี

$$p(B = r) = \frac{4}{10}$$
$$p(B = b) = \frac{6}{10}$$

*** ค่าของความน่าจะเป็นต้องอยู่ในช่วง [0, 1]

ตัวอย่างคำถามที่เกิดขึ้น

- ความน่าจะเป็นที่จะหยิบได้ Apple คือเท่าใด
- ถ้าเราหยิบได้ Orange ความน่าจะเป็นที่หยิบจากกล่องสีน้ำเงิน (Blue) คือ เท่าใด

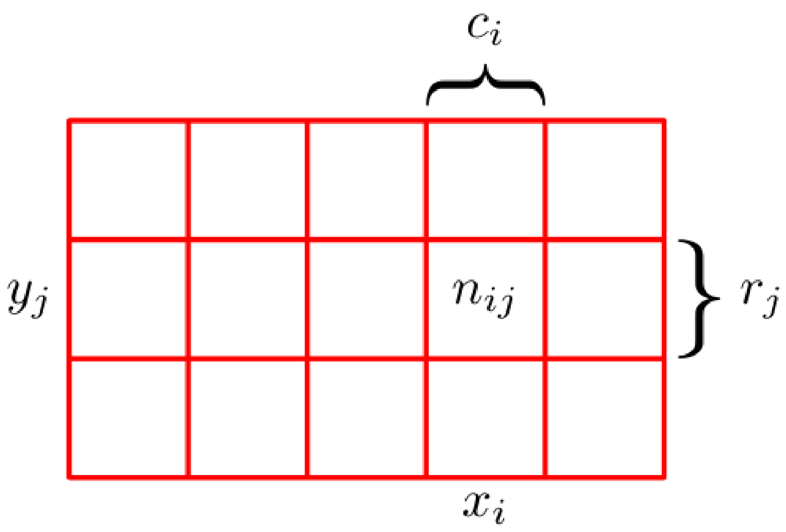
Pattern Recognition - Probability Theory

การเขียนความน่าจะเป็นเมื่อมีเหตุการณ์มากกว่าหนึ่งอย่างเกิดขึ้น
เราจะใช้วิธีการเขียนในรูป **Joint Probability**

$$p(X = x_i, Y = y_i) = \frac{n_{ij}}{N}$$

เราสามารถเขียนความน่าจะเป็นที่เหตุการณ์ X จะมีค่าเท่ากับ x_i โดยไม่สนใจเหตุการณ์อื่นได้ว่า

$$p(X = x_i) = \frac{c_i}{N}$$



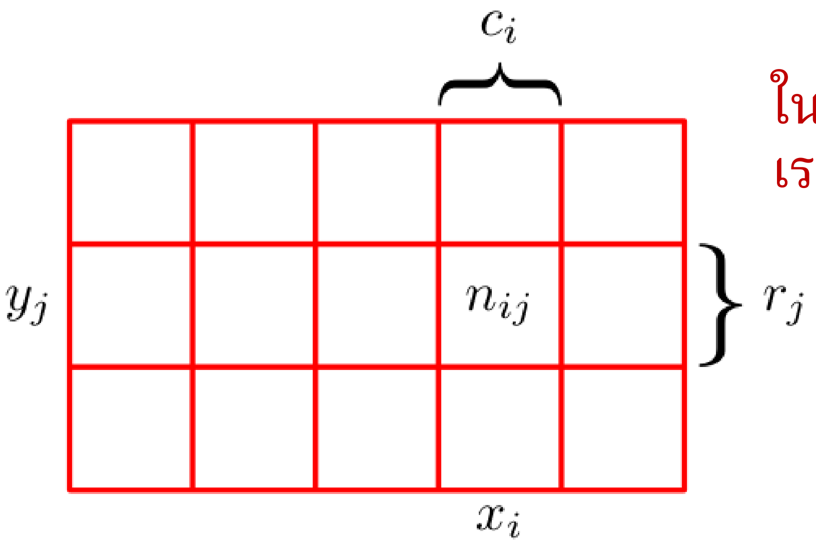
ถ้าพิจารณาจากตารางความน่าจะเป็นแล้วจะเห็นว่า

$$c_i = \sum_j n_{ij}$$

Pattern Recognition - Probability Theory

เมื่อเรานำผลลัพธ์จากสมการทั้งสองก่อนหน้านี้มารวมกันเราจะสามารถ
เขียนความสัมพันธ์ได้ในลักษณะต่อไปนี้ เรียกว่า **Sum Rule**

$$p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_i)$$



ในทางการกลับกันถ้าหากเราสนใจเฉพาะกรณีที่ $(X = x_i)$ เท่านั้น
เราจะเขียนในรูปของ **Conditional Probability**

$$p(Y = y_i | X = x_i) = \frac{n_{ij}}{N}$$

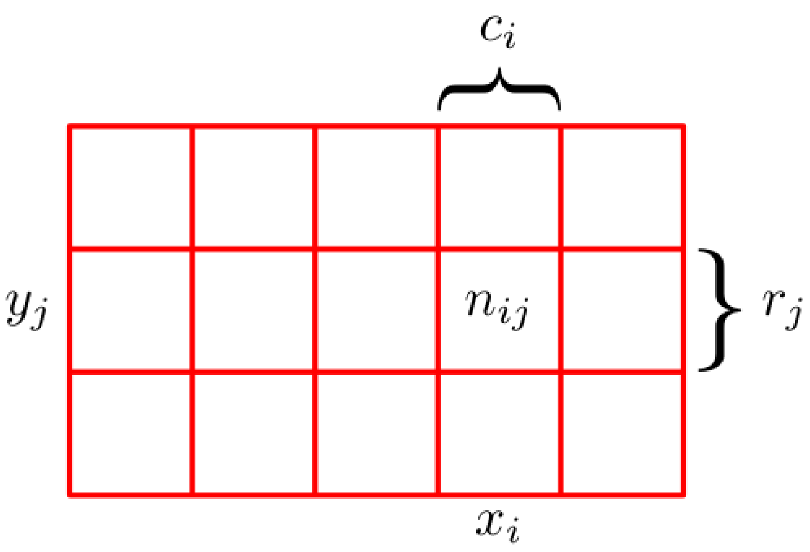
Conditional probability of $Y = y_i$ given $X = x_i$

Pattern Recognition - Probability Theory

จากนั้นเมื่อเราจัดรูปสมการใหม่เราจะเขียนความสัมพันธ์ได้ในรูปต่อไปนี้ ซึ่งเรียกว่า **Product Rule**

$$p(X = x_i, Y = y_i) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$

$$= p(Y = y_i | X = x_i) \cdot p(X = x_i)$$



*** ข้อสังเกต Joint probability มีลักษณะ Symmetric

$$p(X = x_i, Y = y_i) = p(Y = y_i, X = x_i)$$

The Rules of Probability

sum rule $$p(X) = \sum_Y p(X,Y)$$

product rule $$p(X,Y) = p(Y|X)p(X)$$

Pattern Recognition - Probability Theory

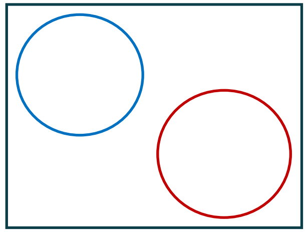
Verbalization

$p(X,Y)$	the probability of X and Y
$p(Y\|X)$	the probability of Y given X
$p(X)$	the probability of X

Pattern Recognition - Probability Theory

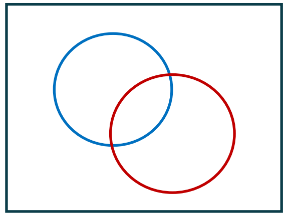
Uncorrelated & Disjoint

Uncorrelated (dis-joint)



$$p(X,Y) = 0$$

Independent



$$p(X,Y) = p(X)p(Y)$$

Example of Independent Events

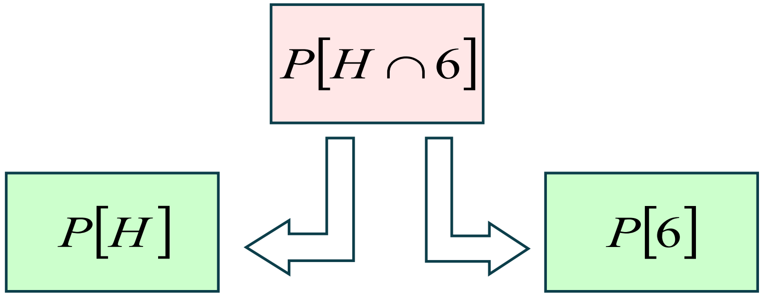
พิจารณาเหตุการณ์ โยนเหรียญได้หัว <u>และ</u> ลูกเต๋าได้ **6** แต้ม

[1] ผลการโยนเหรียญไม่ขึ้นกับผลของลูกเต๋า

[2] ผลของลูกเต๋าไม่ขึ้นกับผลของการโยนเหรียญ

$$P[coin \mid dice] = P[coin] \quad \text{......} \quad [1]$$

$$P[dice \mid coin] = P[dice] \quad \text{......} \quad [2]$$



Pattern Recognition - Probability Theory

เนื่องจาก $p(X,Y) = p(Y,X)$ เมื่อเราใช้ Rules of Probability ที่ผ่านมาจะได้ว่า

$$\boxed{p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}}$$

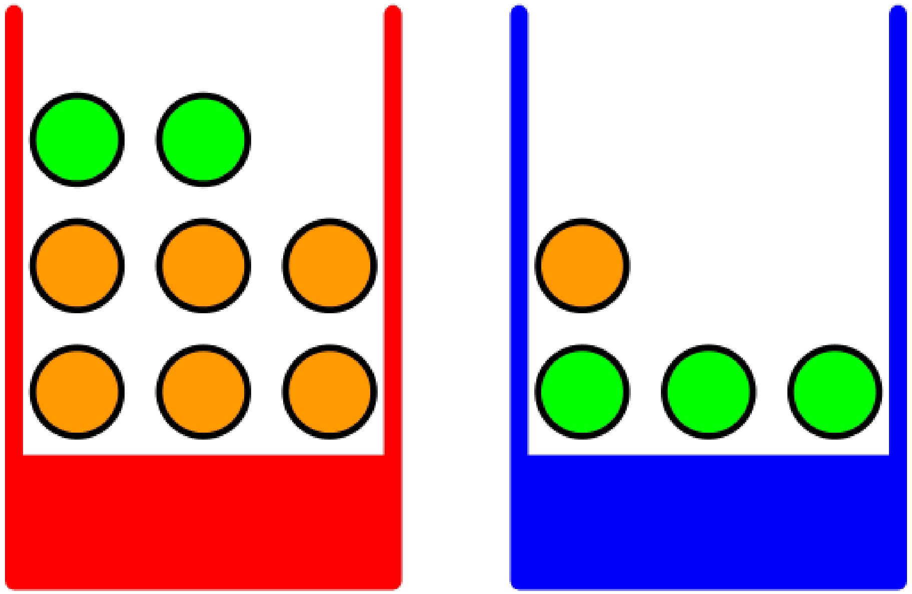
สมการนี้คือ **“Bayes’ theorem”** หรือ **Bayes rule**

เมื่อใช้ Rule of Prob. จัดรูป

$$p(X) = \sum_{Y} p(X|Y)p(Y)$$

Pattern Recognition - Probability Theory

$$
\begin{aligned}
p(B=r) &= 4/10 \\
p(B=b) &= 6/10
\end{aligned}
$$



$$
\begin{aligned}
p(F=a|B=r) &= 1/4 \\
p(F=o|B=r) &= 3/4 \\
p(F=a|B=b) &= 3/4 \\
p(F=o|B=b) &= 1/4
\end{aligned}
$$

Pattern Recognition - Probability Theory

EX. One card is drawn from a deck of cards.

What is the probability that it is a face card, if we know the card is black?

Pattern Recognition - Probability Theory

EX. A bag contains 5 red and 4 white marbles. A marble is drawn from the bag, its color recorded and the marble is returned to the bag. A second marble is then drawn. What is the probability that the first marble is red, and the second marble is white?

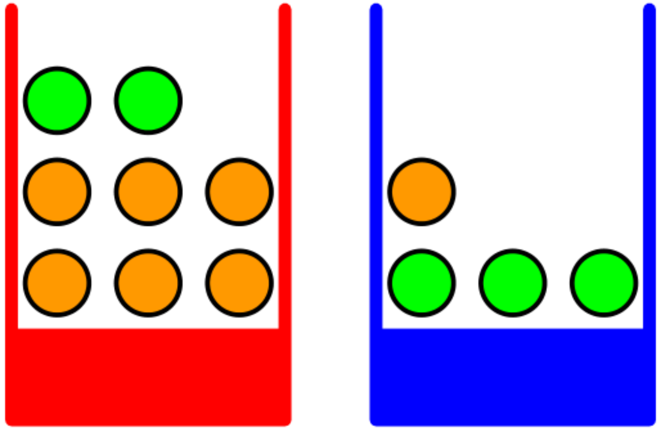
Pattern Recognition - Probability Theory

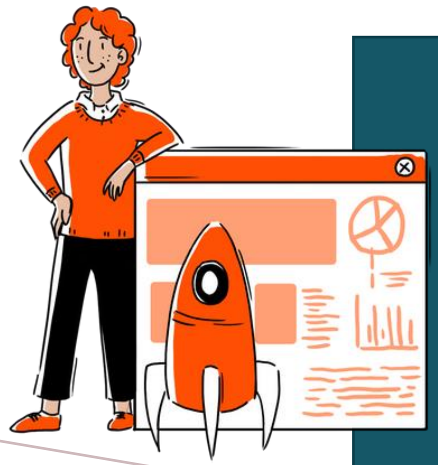
กำหนดให้โอกาสหยิบกล่องแดง 40% และน้ำเงิน 60%

ความน่าจะเป็นที่จะหยิบได้ Apple : $p(F = a) =?$

ความน่าจะเป็นที่จะหยิบได้ Orange : $p(F = o) =?$

เมื่อหยิบส้มมาได้ อยากทราบว่ามีความน่าจะเป็นที่จะมาจากกล่องสีแดงเท่ากับเท่าใด ?





เขียนโปรแกรม (.py) เพื่อหาผลลัพธ์ตามที่กำหนดให้

[1] กำหนดความน่าจะเป็นของการหยิบกล่องแต่ละกล่องได้

[2] กำหนดจำนวนผลไม้ทั้ง 2 ชนิด ในกล่องแต่ละใบได้

[3] ส่ง Program ผ่านทาง GitHub (ให้แยก Folder เป็นวันที่ 2025_11_17)

Bayes Decision Theory

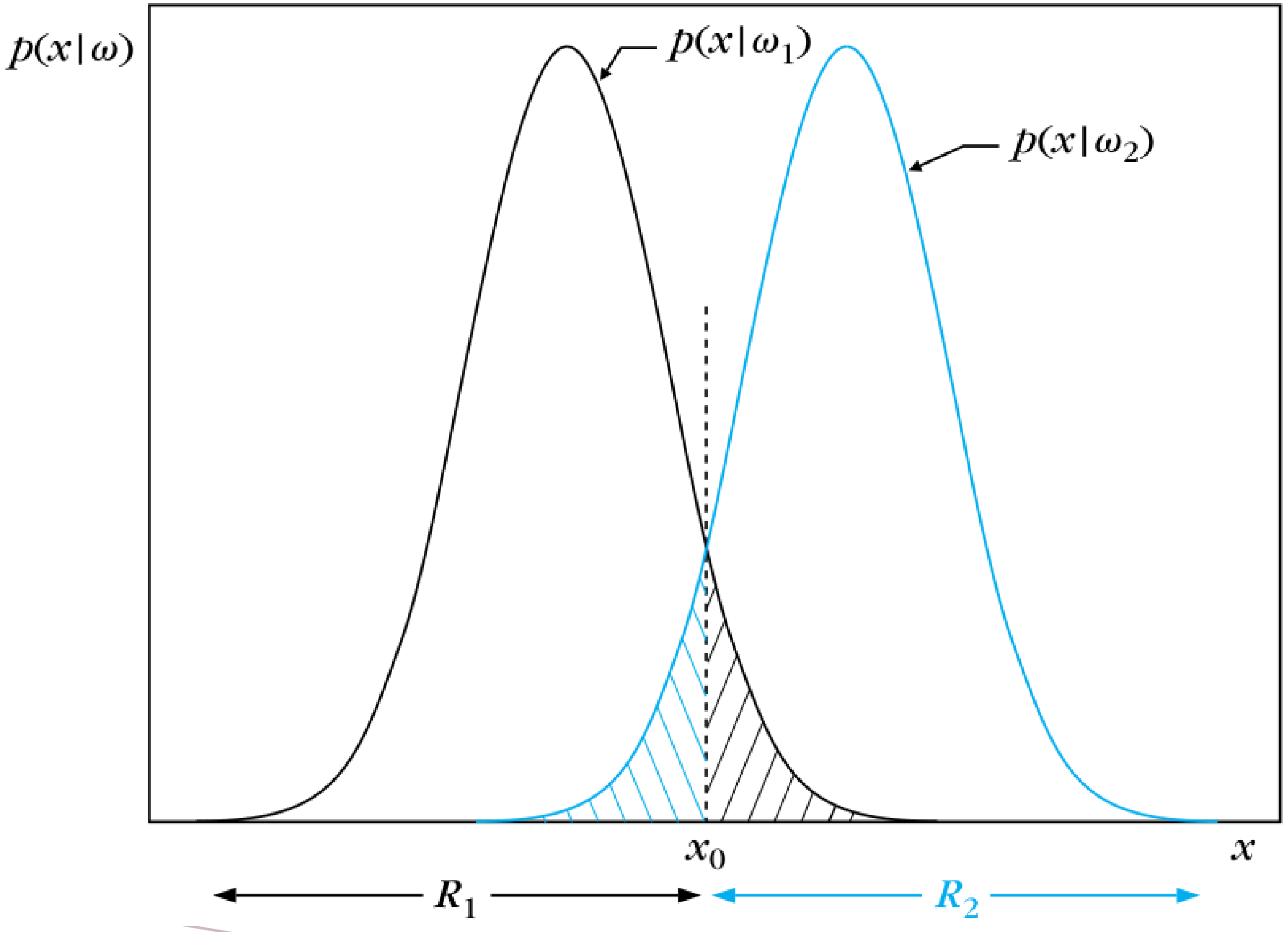
ในการตัดสินคัดแยกข้อมูล (Data Sample) ออกจากกันตามกลุ่ม (Class หรือ Label) จาก Bayes rule เราสามารถเขียนได้ว่า

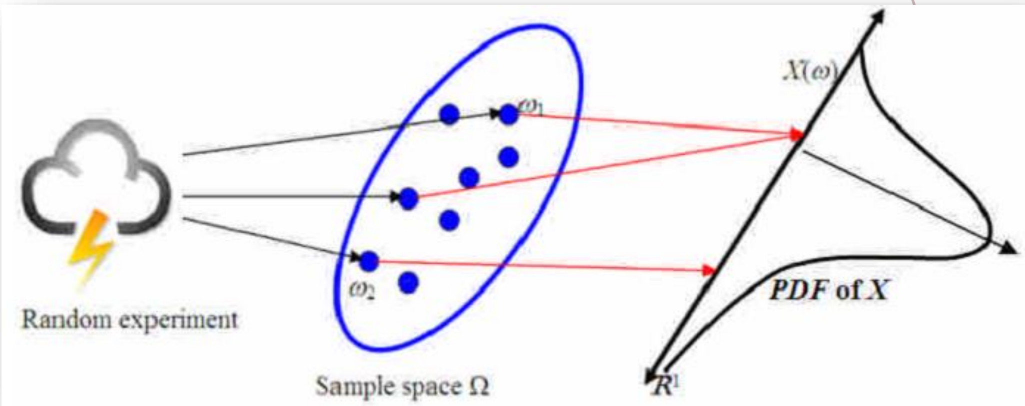
If	$p(\omega_1 \| x) > p(\omega_2 \| x)$	x is classify to class ω_1
If	$p(\omega_1 \| x) < p(\omega_2 \| x)$	x is classify to class ω_2

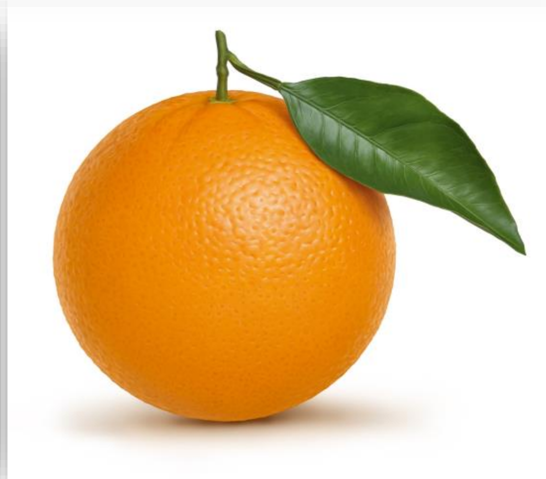
The decision can equivalently be based on the inequalities

$$\boxed{p(x|\omega_1)p(\omega_1) \gtrless p(x|\omega_2)p(\omega_2)}$$

Bayes Decision Theory







Bayes Decision Theory

ตัวอย่าง กรณีทำการแบ่งข้อมูลออกเป็น 2 กลุ่ม (2 Class)

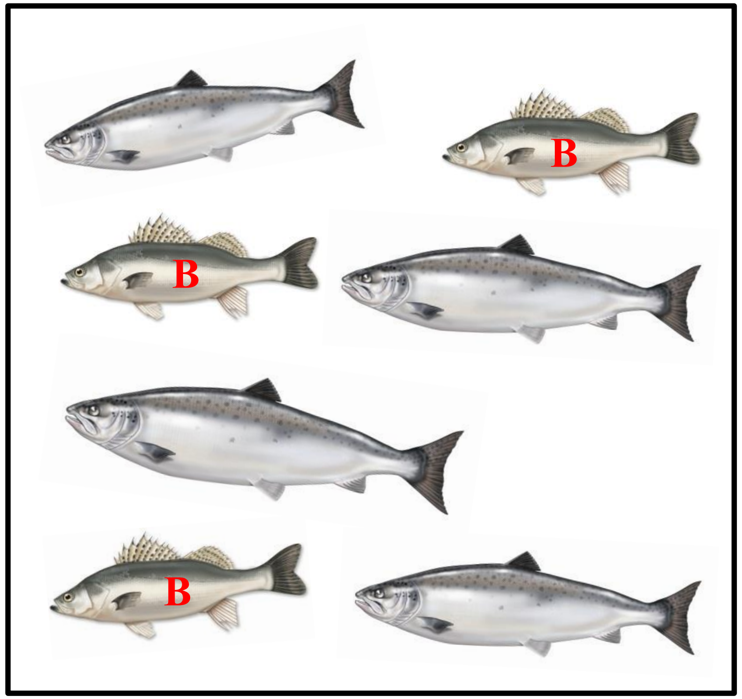
ค่าความน่าจะเป็นที่ข้อมูลจะอยู่ในแต่ละ Class สามารถประมาณค่าได้จากอัตราส่วนระหว่างจำนวนของข้อมูลที่ใช้ในการสอนระบบ (training set) ในแต่ละ Class ต่อจำนวนข้อมูลทั้งหมด

$$P(\omega_1) \approx \frac{N_1}{N}$$

$$P(\omega_2) \approx \frac{N_2}{N}$$

ω_1, ω_2 คือ Class ในแต่ละ Class

N_1, N_2 คือ จำนวนข้อมูลแต่ละ Class



Example

$$P(\text{Salmon}) \approx \frac{N_{\text{Salmon}}}{N} = \frac{4}{7}$$

$$P(\text{SeaBass}) \approx \frac{N_{\text{SeaBass}}}{N} = \frac{3}{7}$$

Bayes Decision Theory

สมมติให้คุณลักษณะ (Feature) ที่ใช้ในการตัดสินใจว่าเป็นปลาชนิดใด คือ **ค่าความยาว** โดยกำหนดให้มีค่าเท่ากับ x จากนั้น เมื่อทำการเขียนให้อยู่ในรูปสมการการตัดสินใจจะได้ว่า

$$\text{If } P(\omega_1 \mid x) > P(\omega_2 \mid x),\ x \text{ is classified to } \omega_1$$

$$\text{If } P(\omega_1 \mid x) < P(\omega_2 \mid x),\ x \text{ is classified to } \omega_2$$

ซึ่งในทางปฏิบัติเราไม่สามารถหาความน่าจะเป็นที่จะตัดสินใจเป็น Class ใด เมื่อรู้ค่าของ Feature ได้ ดังนั้นจึงต้องใช้ความสัมพันธ์จากสมการ Bayes เพื่อจัดรูปใหม่

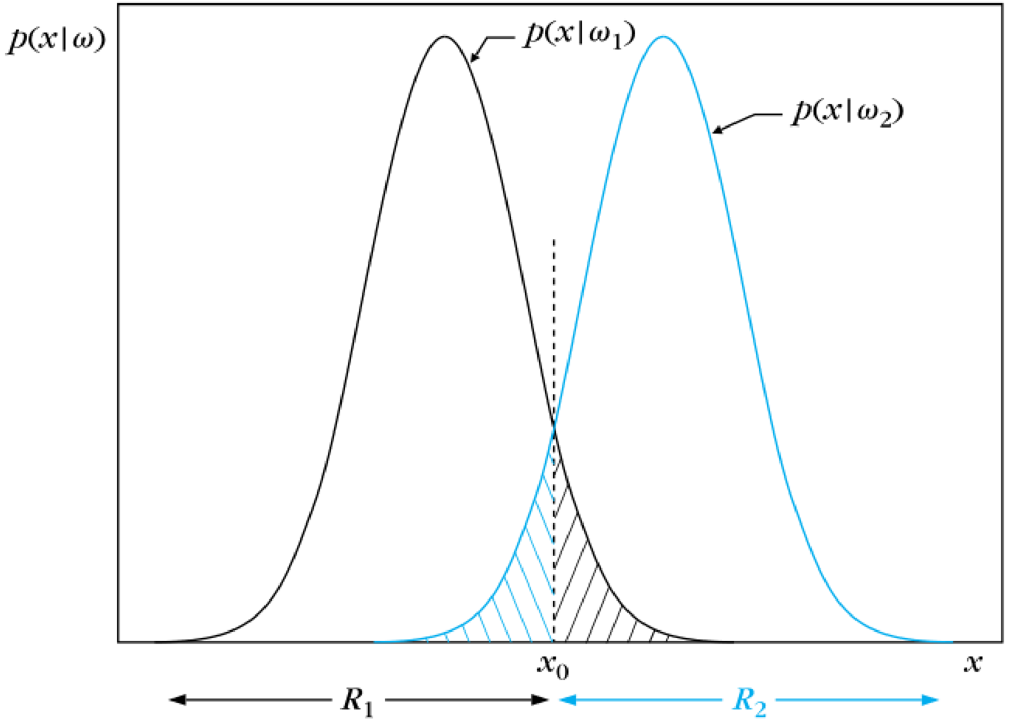
$$\boxed{P(\omega_i \mid x) = \frac{P(x \mid \omega_i)P(\omega_i)}{P(x)}}$$

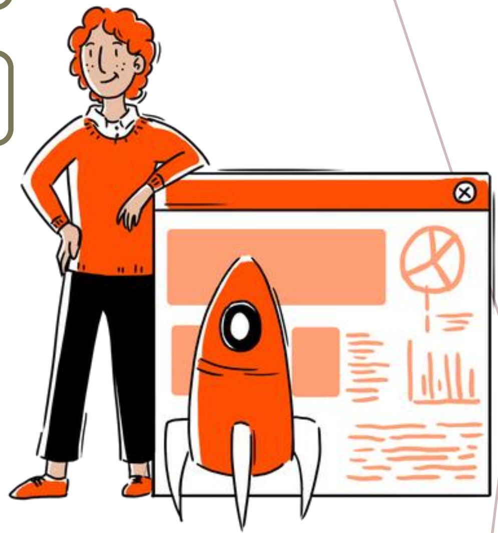
$$\boxed{P(\omega_i \mid x) = \frac{P(x \mid \omega_i)P(\omega_i)}{P(x)}}$$

Bayes Decision Theory

$$P(x \mid \text{Salmon})P(\text{Salmon}) > P(x \mid \text{Seabass})P(\text{Seabass}) \Rightarrow \boxed{\text{Salmon}}$$

$$P(x \mid \text{Salmon})P(\text{Salmon}) < P(x \mid \text{Seabass})P(\text{Seabass}) \Rightarrow \boxed{\text{Seabass}}$$





ฝึกสร้างโปรแกรมต่อไปนี้

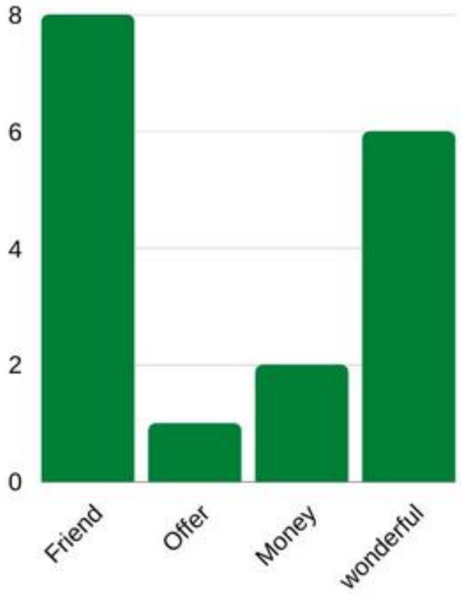
1.) กำหนดจำนวนปลาทั้ง 2 ชนิด

2.) เขียนโปรแกรมใส่ข้อมูลความยาวของปลาทั้ง 2 ชนิดแต่ละตัว

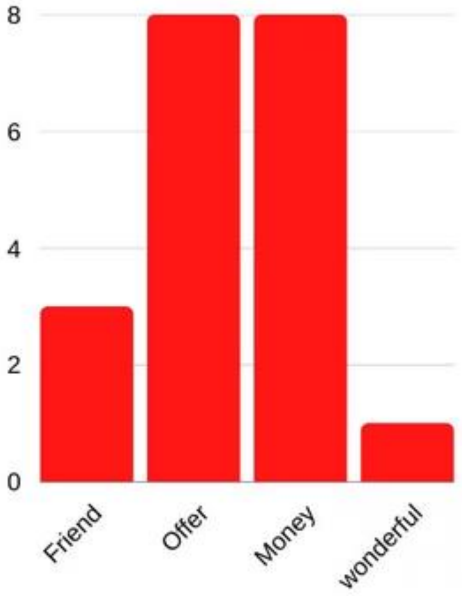
3.) Plot กราฟ Histogram ความยาวของปลาทั้ง 2 ชนิด

4.) เขียนโปรแกรมรับค่า “ความยาว” จากนั้นให้โปรแกรมตัดสินใจว่าเป็นปลาชนิดใด โดยพิจารณาจาก Bayes

Review



Not SPAM



SPAM

$$P(\text{Friend} \mid \text{NOT-SPAM}) = \frac{8}{17} = 0.47 \qquad P(\text{Friend} \mid \text{SPAM}) = \frac{3}{20} = 0.15$$

$$P(\text{Offer} \mid \text{NOT-SPAM}) = \frac{1}{17} = 0.058 \qquad P(\text{Offer} \mid \text{SPAM}) = \frac{8}{20} = 0.40$$

$$P(\text{Money} \mid \text{NOT-SPAM}) = \frac{2}{17} = 0.11 \qquad P(\text{Money} \mid \text{SPAM}) = \frac{8}{20} = 0.40$$

$$P(\text{Wonderful} \mid \text{NOT-SPAM}) = \frac{6}{17} = 0.35 \qquad P(\text{Wonderful} \mid \text{SPAM}) = \frac{1}{20} = 0.05$$

$$P(\text{NOT-SPAM}) = \frac{8}{12} = 0.666 \qquad P(\text{SPAM}) = \frac{4}{12} = 0.333$$

$$P(\text{Friend}|\text{NOT-SPAM}) = \frac{\text{Count of word Friends in Not SPAM corpus}}{\text{Total words in NOT-SPAM Corpus}}$$

Review

Suppose we get an Email: "**Offer Money**" and based on our previously calculated probabilities we need to classify it as SPAM or NOT-SPAM.

$$P(\text{NOT-SPAM} \mid \text{Offer}, \text{Money}) \propto P(\text{NOT-SPAM}) \times P(\text{Offer} \mid \text{NOT-SPAM}) \times P(\text{Money} \mid \text{NOT-SPAM})$$

$$P(\text{NOT-SPAM} \mid \text{Offer}, \text{Money}) \propto 0.666 \times 0.058 \times 0.11$$

$$P(\text{NOT-SPAM} \mid \text{Offer}, \text{Money}) \propto 0.00424$$

$$P(\text{SPAM} \mid \text{Offer}, \text{Money}) \propto P(\text{SPAM}) \times P(\text{Offer} \mid \text{SPAM}) \times P(\text{Money} \mid \text{SPAM})$$

$$P(\text{SPAM} \mid \text{Offer}, \text{Money}) \propto 0.333 \times 0.40 \times 0.40$$

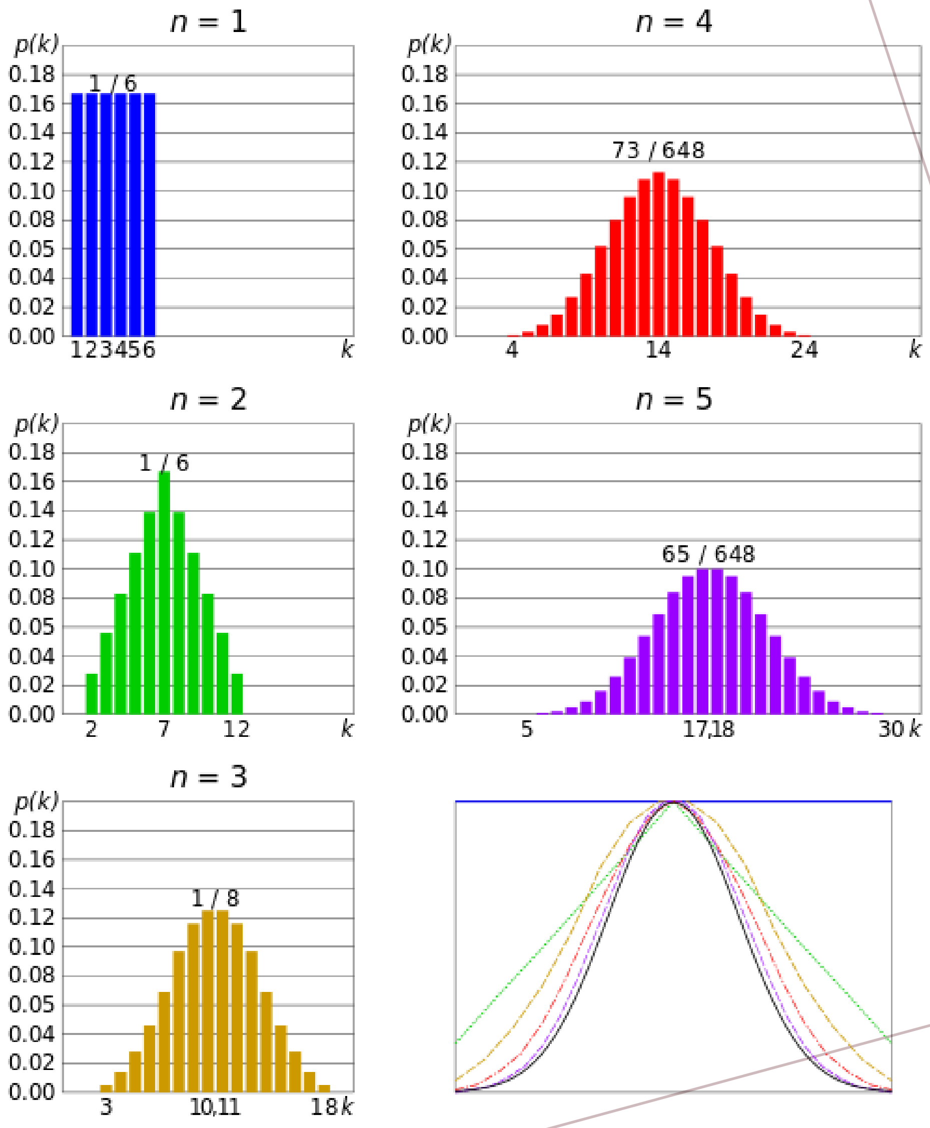
$$P(\text{SPAM} \mid \text{Offer}, \text{Money}) \propto 0.0532$$

Central Limit Theorem

เมื่อมีตัวแปรสุ่มหลายตัวแปรที่มีคุณสมบัติอิสระต่อกัน (Independent) จากทฤษฎีของ Central Limit Theorem จะทำให้เราสามารถประมาณคุณลักษณะของระบบที่เราสนใจด้วยการกระจายตัวแบบ Gaussian Distribution ได้

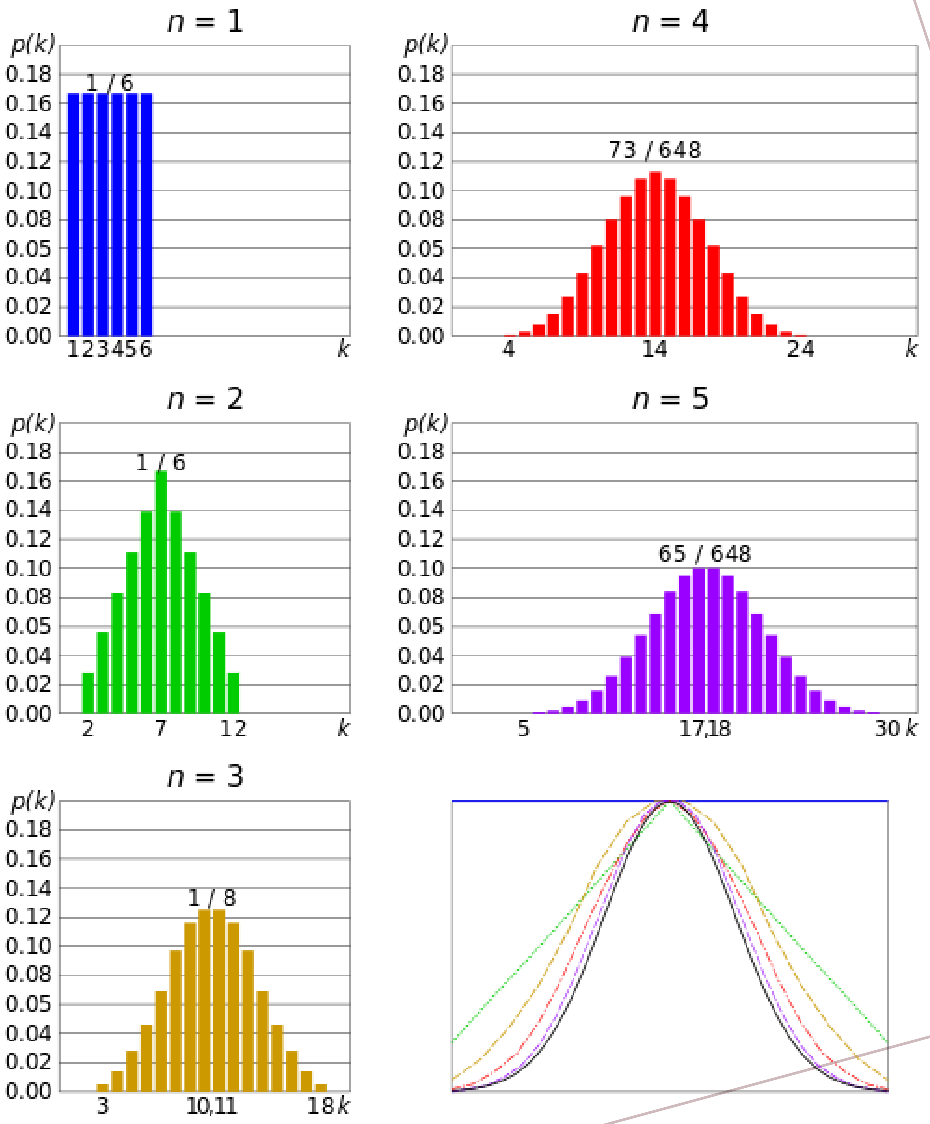
Key Components of the Central Limit Theorem

1. **Sampling is successive.** This means some sample units are common with sample units selected on previous occasions.
2. **Sampling is random.** All samples must be selected at random so that they have the same statistical possibility of being selected.
3. **Samples should be independent.** The selections or results from one sample should have no bearing on future samples or other sample results.
4. **Samples should be limited.** It's often cited that a sample should be no more than 10% of a population if sampling is done without replacement. In general, larger population sizes warrant the use of larger sample sizes.
5. **Sample size is increasing.** The central limit theorem is relevant as more samples are selected.



Central Limit Theorem

$$\bar{X}_n \xrightarrow[n\to\infty]{} N\left(\mu, \frac{\sigma^2}{n}\right)$$



Sample size…Why 30 ?

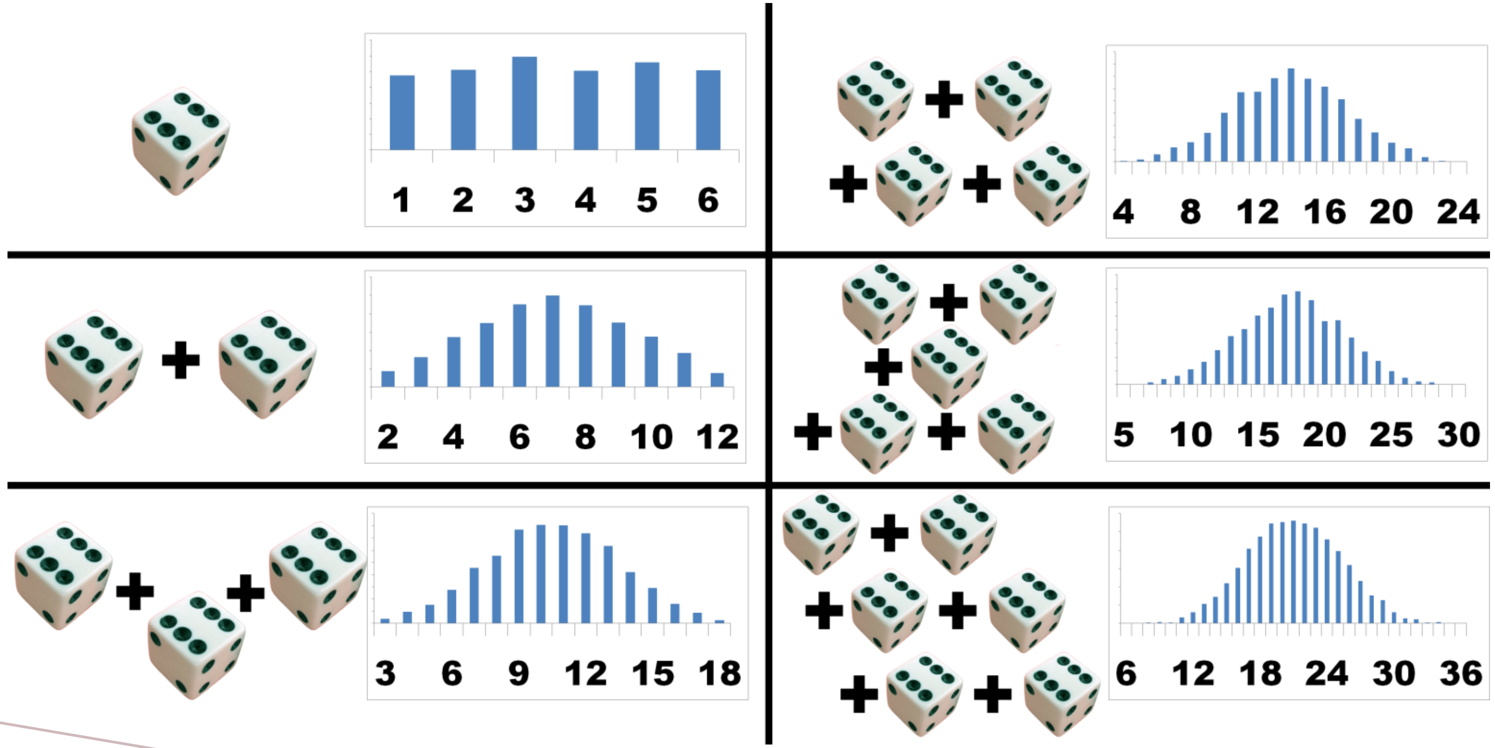
You need at least 30 data sample before you can reasonably expect an analysis based upon the normal distribution to be valid.

This is a crude rule of thumb that suggest that the normal distribution could be a reasonable approximation if you want to make inferences about the mean.

Workshop

Central Limit Theorem

ให้เขียนโปรแกรมทดสอบทฤษฎี Central Limit Theorem ว่าเป็นไปอย่างถูกต้องหรือไม่ โดยกำหนดให้ใช้การสุ่มแบบ Uniform โดยมีค่าอยู่ระหว่าง [1 – 6] แทนหน้าลูกเต๋า และทดสอบผลลัพธ์ที่ได้จากการนำค่าที่ออกมาบวกกัน




KEEP IT UP