Probability and Statistics

Eng. Mahmoud Ouf

Lecture 01

Set theory

The set theory involves mathematical operations at a set level.

It is thought as the basic fundamental group of theorems that governs the rest of mathematics.

For our purpose, we use the set theory in order to manipulate groups of elements.

A **set** is a collection of distinct elements.

Note: the element in a set, can't be repeated

The **magnitude** of a set is the number of elements in the set.

The concept of an empty set exists and is denoted by the {} or φ character .

This empty set is said to have a magnitude of 0.

<u>Set theory</u>

Operation in Sets

Assume set X for best friends, set Y for Soccer Team, set Z for Tennis Team

X ={alex, blair, casey, drew, erin, francis, glen, hunter, ira, jade}

Y ={alex, casey, drew, hunter}

Z ={casey, drew, jade}

element of ($a \in A$): this mean that the element (a) is part of the set A

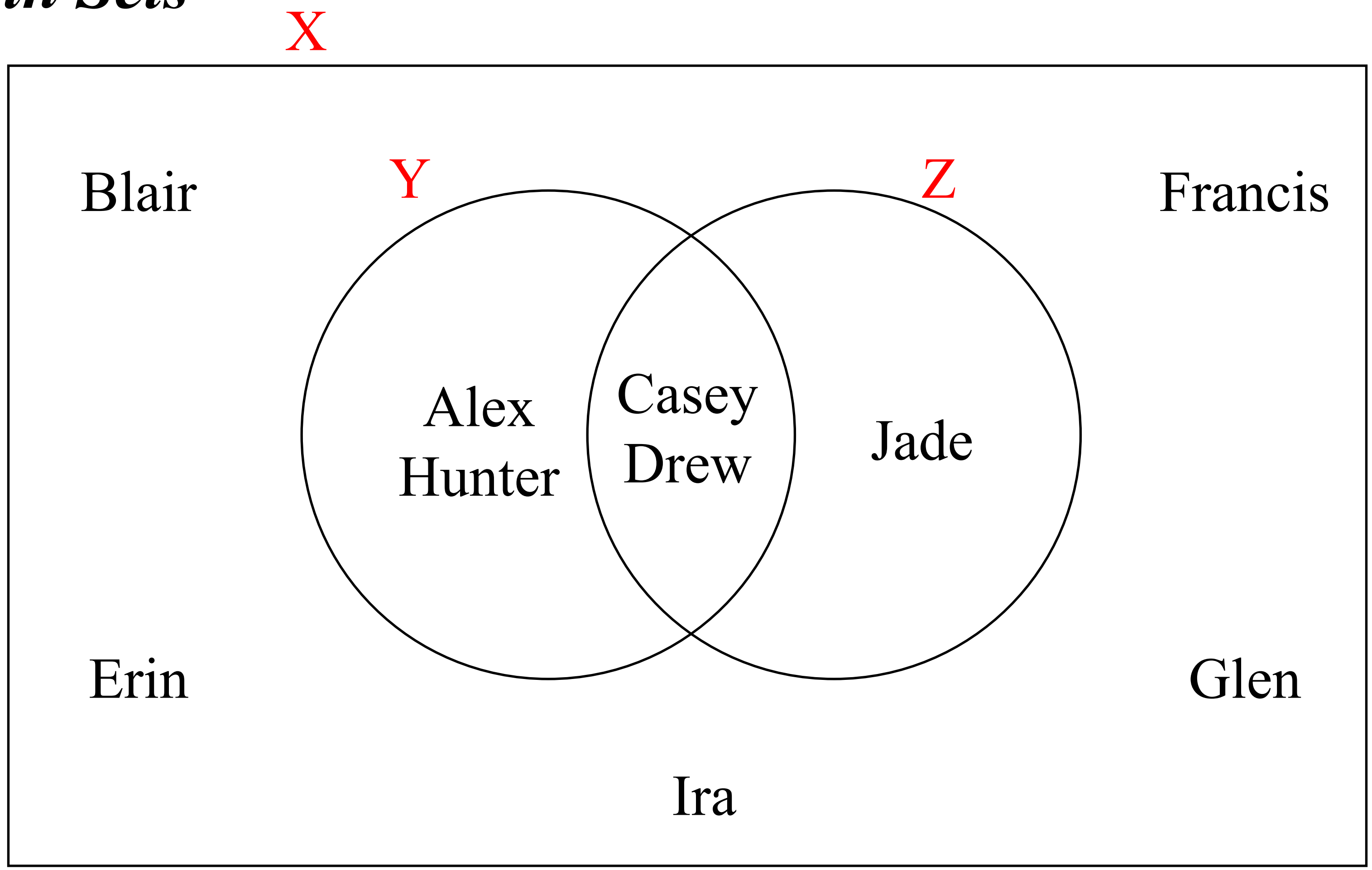
$$\boldsymbol{erin \in X}$$

not element of ($a \notin A$): this mean that the element (a) is not part of the set A

$$\boldsymbol{jan \notin X}$$

Set theory

Operation in Sets



Set theory

Operation in Sets

intersection ($A \cap B$): elements that belong to set A **and** set B

$$Y \cap Z = \{casey, drew\}$$

union ($A \cup B$): elements that belongs to set A **or** set B

$$Y \cup Z = \{alex, casey, drew, hunter, jade\}$$

proper subset/strict subset ($A \subset B$): all elements of A are part of B, and A has fewer elements than B

$$Y \subset X$$

subset ($A \subseteq B$): all elements of A are part of B, and A has the same elements or fewer than B

$$X \subseteq X$$

Set theory

Operation in Sets

not subset ($A \not\subset B$): all elements of A are not part of B

$$Z \not\subset Y$$

Note: if($A \subseteq B$), then A is subset of B and B is superset of A

equality (A = B): both sets A and B have the same elements (identical)

complement (A^c or A'): All elements that doesn't belong to A

$$Y^c = \{blair, erin, francis, glen, ira, jade\}$$

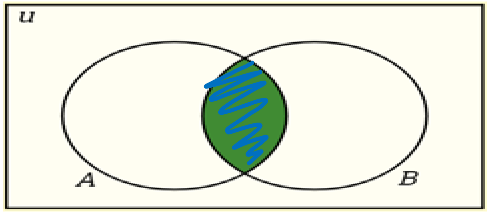
relative complement (A-B): elements that belong to A and not B

$$Y\text{-}Z = \{alex, hunter\}$$

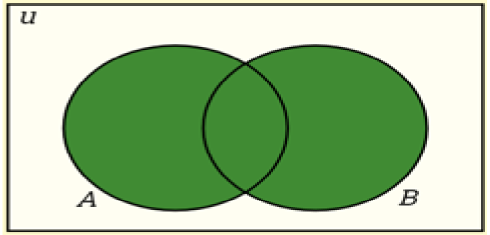
Set theory (using Venn Diagram)

Operation in Sets

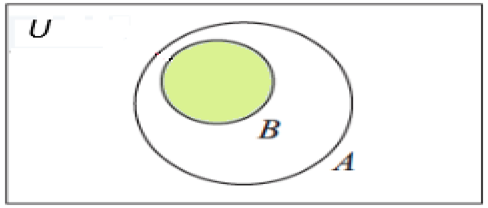
intersection (A∩B):



union (A∪B):



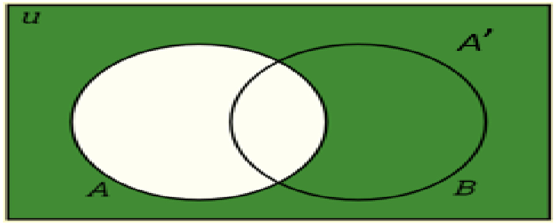
proper subset/strict subset (B⊂A):



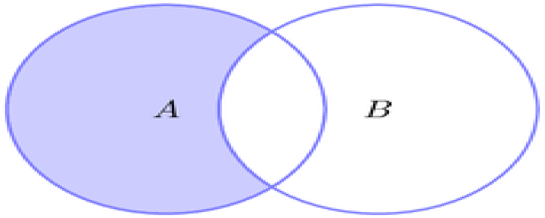
Set theory (using Venn Diagram)

Operation in Sets

complement (A^c or A’):



relative complement (A-B):



Set theory

Properties of Sets

Commutative law: $A \cup B = B \cup A$

$A \cap B = B \cap A$

Associative law: $(A \cup B) \cup C = A \cup (B \cup C)$

$(A \cap B) \cap C = A \cap (B \cap C)$

Distributive law: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

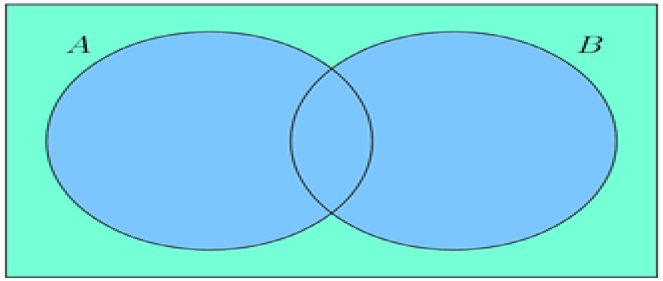
$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

Set theory

DeMorgan's laws

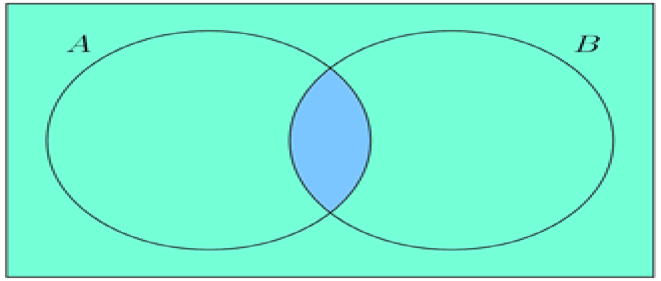
$(A \cup B)^c = A^c \cap B^c$

NOT (A or B) ➔ NOT A AND NOT B



$(A \cap B)^c = A^c \cup B^c$

NOT (A and B) ➔ NOT A OR NOT B



Probability

Probability is the measure of the likelihood that an event will occur.

Probability is quantified as a number between 0 and 1, where, 0 indicates impossibility and 1 indicates certainty.

The higher the probability of an event, the more likely it is that the event will occur.

Sample Space (S)

Sample space is a set of all possible outcomes from the experiment.

A sample space is *discrete* if it consists of a finite or countable infinite set of outcomes.

A sample space is *continuous* if it contains an interval (either finite or infinite) of real numbers.

Probability

Example 1: when throwing a dice

S = {1, 2, 3, 4, 5, 6}

Example 2: when throwing a coin 2 times, what will be the sample space?

S = {HH, HT, TH, TT}

Example 3: We have a box contains 3 balls (Red, Green, Blue), we need to draw 2 balls, define the sample space

Without replacement:

S = {RG, RB, GR, GB, BR, BG}

With replacement:

S = {RR, RB, RG, BR, BB, BG, GR, GB, GG}

Probability

Note:

Sampling without replacement is more common for industrial applications.

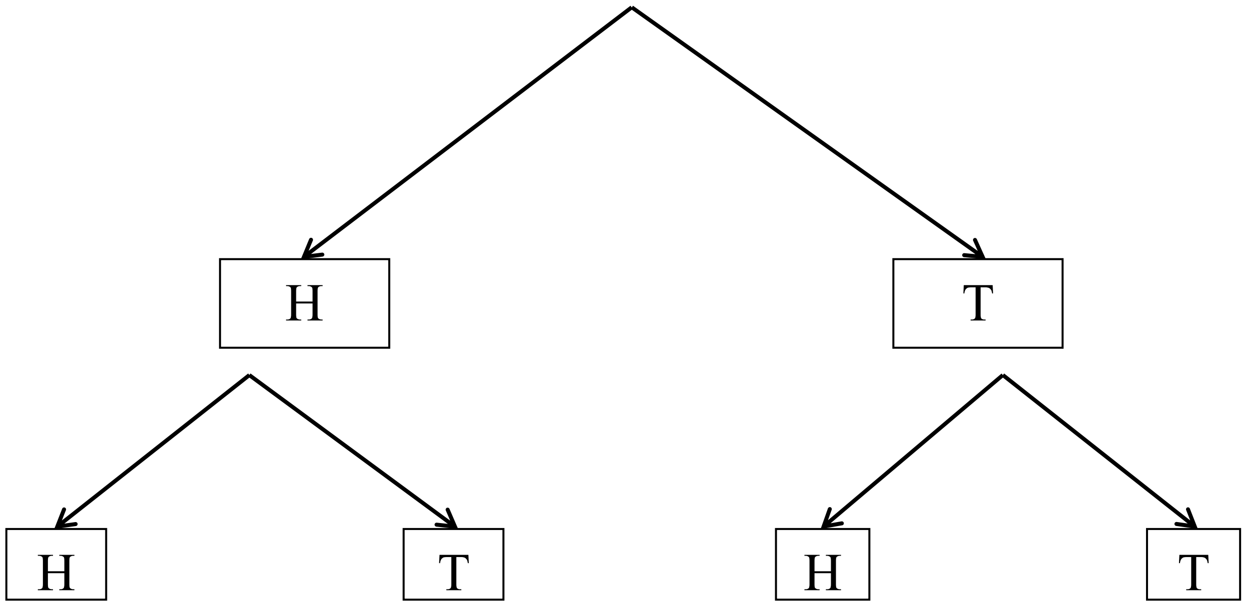
The above sample space, we respect the order of selection.

If the order of the selected item is not effective, then we will get another sample space. Example the sample space of the above case without replacement will be: S = {{R, G},{R, B},{G, B}}

Sample spaces can also be described graphically with tree diagrams.

Example throwing a coin 2 times

Probability



Events

An event is a subset of the sample space of a random experiment.

Example: getting 2 head when throwing a coin twice {HH}

Probability

Because events are subsets, we can use basic set operations such as unions, intersections, and complements to form other events of interest.

Note:

Assume E1 and E2 are 2 events

- If $E1 \cap E2 = \varphi$, the we say that E1 and E2 are **mutually exclusive**
- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
- $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
- $(A \cup B)' = A' \cap B'$ and $(A \cap B)' = A' \cup B'$

Probability

It provides the chance of an Event to occur. The probability value is a value between 0 and 1.

Probability

0 indicates impossible to occur, 1 indicates certainly it will occur. Higher values indicate that the event will more likely occur than lower values.

When the model of equally likely events is assumed, the probabilities are chosen to be equal.

A sample space consists of N possible events that are equally likely; the probability of each event is 1/N

For a discrete sample space, the probability of an event E, denoted as P(E), equals the sum of the probabilities of the elements in E.

If S is the sample space and E is any event in a random experiment,

- $P(S) = 1$
- $P(\varphi) = 0$
- For 2 events E1 , E2 : $P(E1 \cup E2) = P(E1) + P(E2) - P(E1 \cap E2)$

Probability

Example 1:

A bag contains 8 marbles numbered 1 to 8

a. What is the probability of selecting a 2 from the bag?

b. What is the probability of selecting an odd number?

c. What is the probability of selecting a number greater than 6?

Solution

S = {1, 2, 3, 4, 5, 6, 7, 8}

a) E = {2}

P(E) = 1 / 8

b) E = {1, 3, 5, 7}

P(E) = 4 / 8 = 1 / 2

c) E = {7, 8}

P(E) = 2 / 8 = 1 / 4

Probability

Example 2:

If you roll 2 dice at the same time,

a) what is the probability the sum is 6 or a pair of odd numbers?

b) Their sum is equal to 1

Solution

$$S = \{ \begin{array}{l} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6) \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6) \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6) \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \end{array} \}$$

Probability

a) $E1 = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$

$E2 = \{(1,1), (1,3), (1,5), (3,1), (3,3), (3,5), (5,1), (5,3), (5,5)\}$

$E1 \cap E2 = \{(1,5), (3,3), (5,1)\}$

$P(E1) = 5 / 36$

$P(E2) = 9 / 36$

$P(E1 \cap E2) = 3 / 36$

$P(E1 \cup E2) = P(E1) + P(E2) - P(E1 \cap E2)$

$= 5/36 + 9/36 - 3/36$

$= 11/36$

b) $E = \{\}$

$P(E) = 0 / 36$

$= 0$

Probability

Example 3

A jar contains 3 red marbles, 7 green marbles and 10 white marbles. If a marble is drawn from the jar at random,

a) what is the probability that this marble is white?

b) What is the probability that this marble is not green?

Solution

a) E (white) = 10

P(Ew) = 10 / 20

= 0.5

Probability

b) E (green) = 7

$P(Eg) = 7 / 20$

$P(\text{Not Green}) = 1 - P(Eg)$

$P(\text{Not Green}) = 1 - 7/20$

$= 13 / 20$

Another Solution:

Probability not Green = Probability White or Red

$P(\text{Not Green}) = P(Ew \cup ER)$

$P(Ew) = 10 / 20$

$P(ER) = 3 / 20$

$P(Ew \cap ER) = 0$ they are mutual exculsive

$P(Ew \cup ER) = P(Ew) + P(ER) - P(Ew \cap ER)$

$= 10/20 + 3/20 - 0$

$= 13 / 20$

Probability

Example 4

A pack contains 4 blue, 2 red and 3 black pens. If a pen is drawn at random from the pack, replaced and the process repeated 2 more times, what is the probability of drawing 2 blue pens and 1 black pen with this order?

Solution

P(EBlue) = 4/9

P(EBlack) = 3/9

P(E2Blue, E1Black) = P(EBlue) * P(EBlue) * P(EBlack)

= 4/9 * 4/9 * 3/9

= 48/729

= 16/243

Counting Probability

In a Sample Space S, if the probability of an event A is the ratio of the numbers of outcomes in A to total number of outcomes in S

In several cases, the number of outcomes in S is too large.

So, we need a method to determine the number of outcome in a simple way

Multiplication rules:

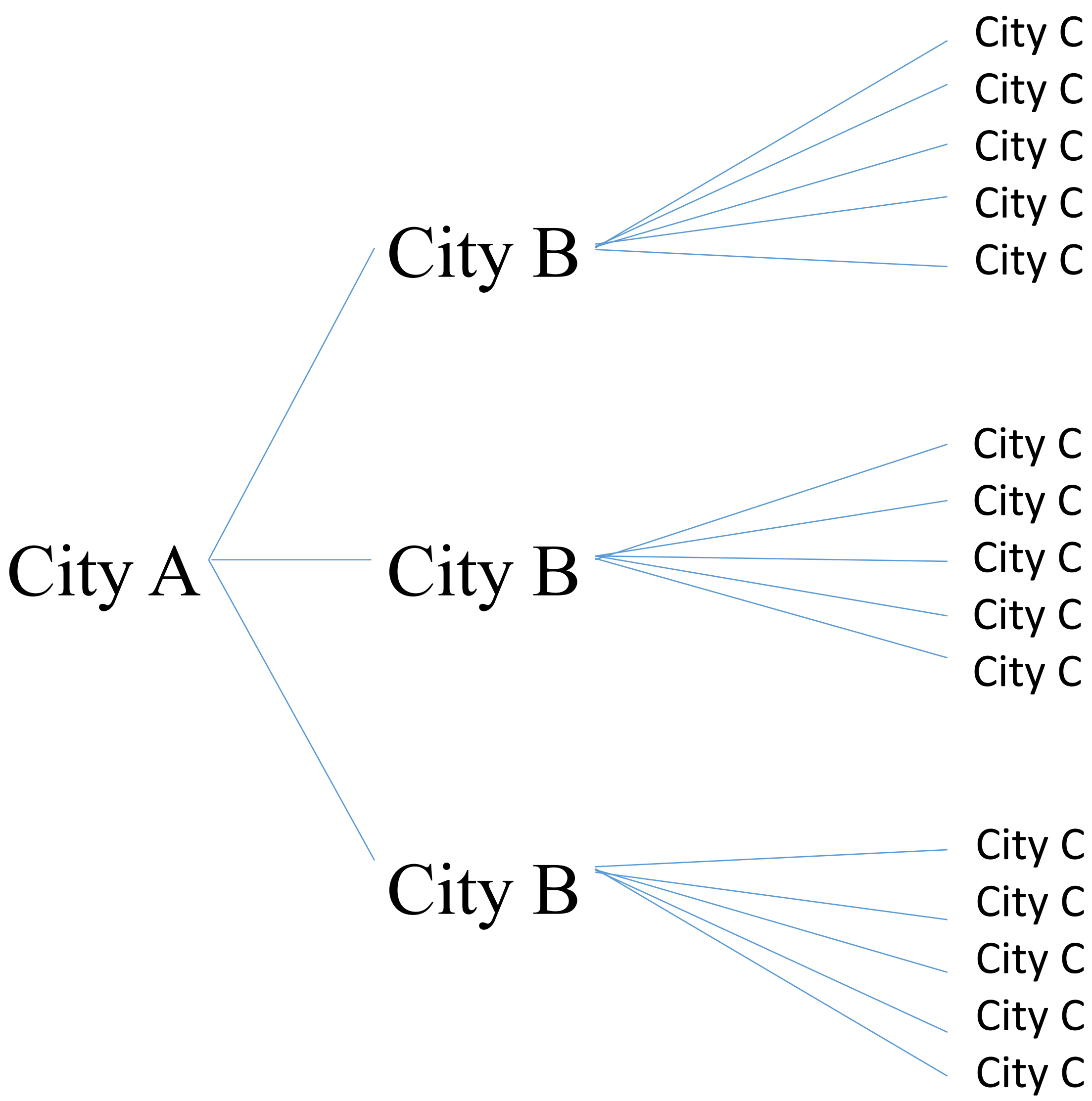
Ex: 1) Route between cities:

Suppose there are 3 routes between city A and city B, and 5 routes from city B to city C

The total number of route that we can get from city A to city C is:

Number of routes = 3 * 5 = 15 routes

Counting Probability



Counting Probability

Ex: 2) Experiment in two parts:

Suppose there are is an experiment performed in 2 parts consecutive , the first part has m possible outcome, the second part has n possible outcome

The total number of outcomes = m * n

If m = 3, n = 2

Number of outcomes = 3 * 2 = 6

Ex: 3) Rolling a dice 2 times (rolling 2 dices one time)

each dice has 6 outcome

Total number of outcomes = 6 * 6 = 36

Ex: 4) Tossing a coin 6 times, each toss has 2 outcome (H, T)

Total number of outcomes = $2 * 2 * 2 * 2 * 2 * 2 = 2^6$

Conditional Probability

The conditional probability of an event B given an event A, means that the probability of event B in the condition that event A has occurred. It is denoted as:

$P(B|A) = P(B \cap A) / P(A)$

$P(B \cap A) = P(B|A) * P(A)$ "multiplication rule"

Example 1:

In an exam, two reasoning problems, 1 and 2, are asked.

35% students solved problem 1 and 15% students solved both the problems.

How many students who solved the first problem will also solve the second one?

Conditional Probability

Solution

Probability of student solving problem 1, $P(1) = 0.35$

Probability of student solving both problem $P(1 \cap 2) = 0.15$

Probability of solving 2 if 1 is solved $P(2|1)$

$P(2|1) = (P(1 \cap 2))/(P(1))$

$= 0.15 / 0.35$

$= 3 / 7$

Conditional Probability

Example 2

Out of 50 people surveyed in a study, 35 smoke in which there are 20 males. What is the probability the if the person surveyed is a smoker then he is a male?

Solution:

Let A the event of being smoker

Let B the event of being male

$P(A \cap B) = 20 / 50$

$P(A) = 35 / 50$

$P(B|A) = P(A \cap B) / P(A)$

$= (20/50) / (35/50)$

$= 4 / 7$

Conditional Probability

Example 3

The probability of raining on Sunday is 0.07.

If today is Sunday then find the probability of rain today.

Solution

Let A the event of Raining

Let B the event of Sunday

$P(A \cap B) = 0.07$

$P(B) = 1/7$

$P(A|B) = P(A \cap B) / P(B)$

$= 0.07 / (1/7)$

$= 0.49$

Conditional Probability

Example 4

What is the probability that the total of two dice will be greater than 9, given that the first die is a 5?

Solution

Let A = first die is 5

Let B = total of two dice is greater than 9

$P(A) = 6 / 36 = 1/6$

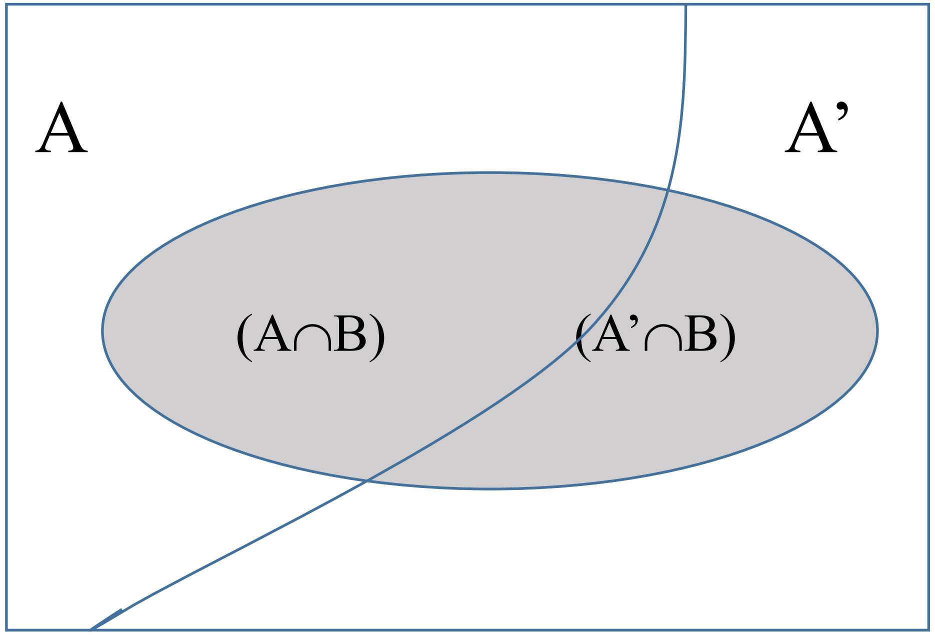
$P(A \cap B) = 2 / 36 = 1 / 18$

$P(B|A) = P(A \cap B) / P(A)$

$= (1/18) / (1/6)$

$= 1 / 3$

<u>Total Probability Rule:</u>



$P(B) = P(B \cap A) + P(B \cap A')$ “A and A’ are mutual exclusive”

$= P(B|A) * P(A) + P(B|A') * P(A')$

Total Probability Rule:

Example:

suppose that in semiconductor manufacturing the probability is 0.10 that a chip that is subjected to high levels of contamination during manufacturing causes a product failure.

The probability is 0.005 that a chip that is not subjected to high contamination levels during manufacturing causes a product failure.

In a particular production run, 20% of the chips are subject to high levels of contamination.

What is the probability that a product using one of these chips fails?

Total Probability Rule:

Solution

Let F denote the event that the product Fail

Let H denote the event that the ship is exposed to high level contamination

$P(F|H) = 0.1$

$P(F|H') = 0.005$

$P(H) = 0.2$

$P(H') = 0.8$

$$P(F) = P(F|H) * P(H) + P(F|H') * P(H')$$
$$= (0.1 * 0.2) + (0.005 * 0.8)$$

Total Probability Rule:

To generalize the rule

$S = E_1 \cup E_2 \cup E_3 \cup \dots \cup E_k$ "Exhaustive Events"

$P(B) = P(B \cap E_1) + P(B \cap E_2) + \dots + P(B \cap E_k)$

$= P(B|E_1)*P(E_1) + P(B|E_2)*P(E_2) + \dots + P(B|E_k)*P(E_k)$

Example

Continuing with the semiconductor manufacturing example, assume the following probabilities for product failure subject to levels of contamination in manufacturing:

Probability of Failure	Level of Contamination
0.1	High
0.01	Medium
0.001	Low

Total Probability Rule:

In a particular production run,

20% of the chips are subjected to high levels of contamination,

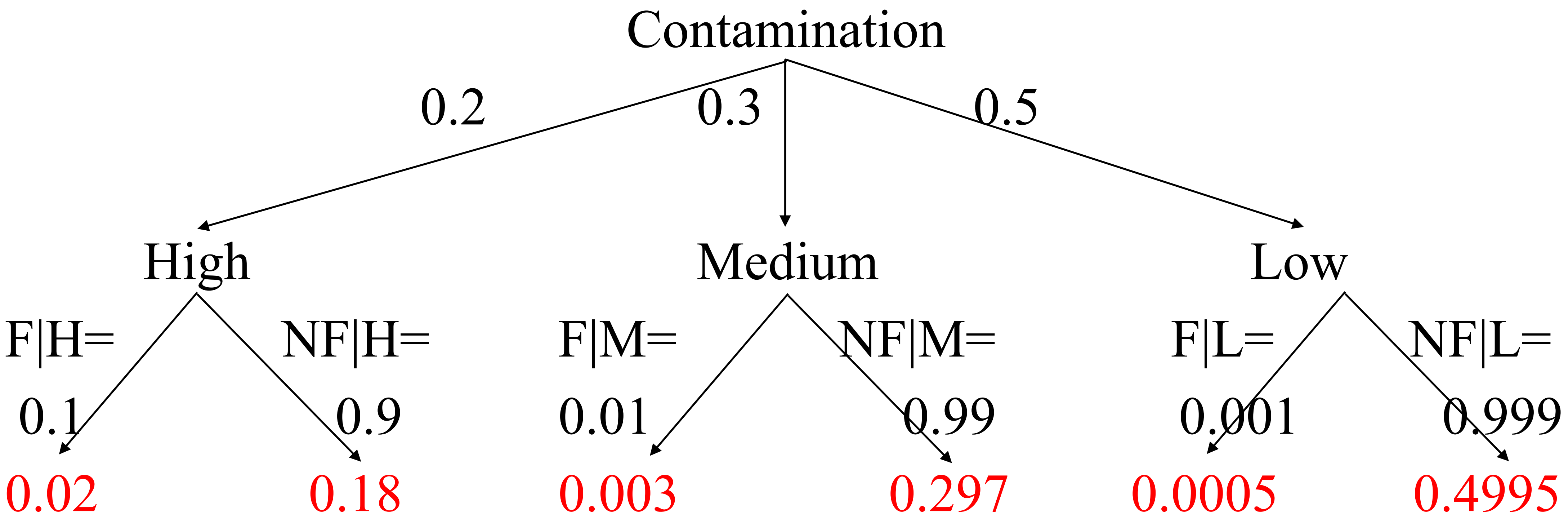
30% to medium levels of contamination,

and 50% to low levels of contamination.

What is the probability that a product using one of these chips fails?

Solution

Total Probability Rule:



P(F) = 0.02 + 0.003 + 0.0005 = 0.0235

Total Probability Rule:

H denote the event that a chip is exposed to high levels of contamination

M denote the event that a chip is exposed to medium levels of contamination

L denote the event that a chip is exposed to low levels of contamination

$$P(F) = P(F|H)P(H) + P(F|M)P(M) + P(F|L)P(L)$$

$$= (0.1 * 0.2) + (0.01 * 0.3) + (0.001 * 0.5)$$

$$= 0.02 + 0.003 + 0.0005$$

$$= 0.0235$$

Independent Events

Event B is said to be independent to Event A, if the Event A has no effect on the Event B.

This mean that the Event B has the same value whatever the value of Event A

Example:

Suppose a day's production of 850 manufactured parts contains 50 parts that do not meet customer requirements.

Suppose two parts are selected from the batch, but the first part is replaced before the second part is selected.

What is the probability that the first part and second part are defective

Independent Events

Solution

Let the first part is defective (denoted as A)

Let the second part is defective (denoted as B)

$P(A) = 50 / 850$

Because the first part is replaced prior to selecting the second part, the batch still contains 850 parts, of which 50 are defective.

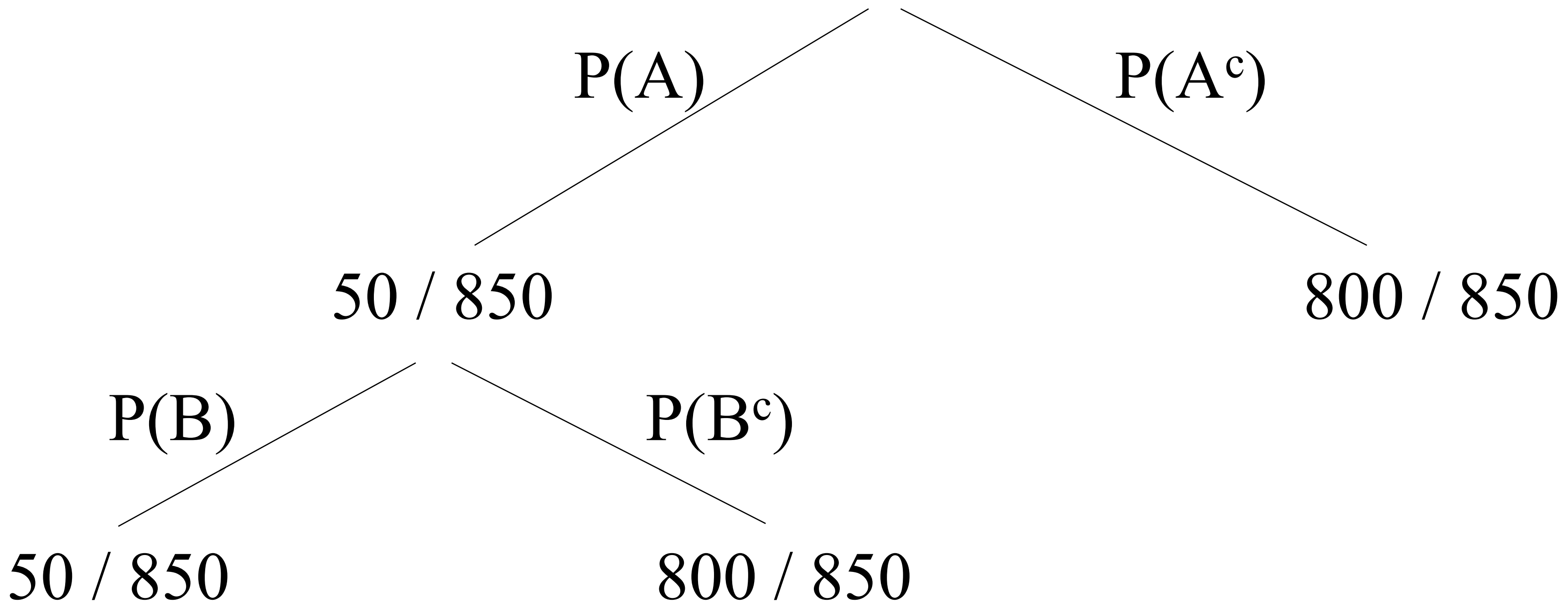
Therefore, the probability of B does not depend on whether or not the first part was defective.

$P(B) = 50 / 850$

$$
\begin{aligned}
P(A \cap B) &= P(A)P(B) \\
&= (50/850)(50/850) \\
&= 0.0035
\end{aligned}
$$

Independent Events

Solution



Rule:

If A and B are two independent Events, then:

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

Independent Events

Example:

Suppose that two machines 1 and 2 in a factory are operated independently of each other.

Let A be the event that machine 1 will become inoperative during a given 8-hour period,

Let B be the event that machine 2 will become inoperative during the same period.

Suppose that $P(A) = 1/3$ and $P(B) = 1/4$.

We shall determine the probability that at least one of the machines will become inoperative during the given period.

Independent Events

Solution:

Both event A and event B are independent, machine 1 has no effect on machine 2 and machine 2 has no effect on machine 1.

$$P(A \cap B) = P(A) * P(B)$$
$$= (1/3) * (1/4)$$
$$= (1/12)$$

Probability that at least one of the machine become inoperative $P(A \cup B)$:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= (1/3) + (1/4) - (1/12)$$
$$= 1 / 2$$

Independent Events

Example:

Suppose that a balanced die is rolled.

Let A be the event that an even number is obtained

Let B be the event that one of the numbers 1, 2, 3, or 4 is obtained.

Show that the events A and B are independent.

Solution

$S = \{1, 2, 3, 4, 5, 6\}$

$A = \{2, 4, 6\}$ ➔ $P(A) = 1/2$

$B = \{1, 2, 3, 4\}$ ➔ $P(B) = 2/3$

$A \cap B = \{2, 4\}$ ➔ $P(A \cap B) = 1/3$ → (I)

$P(A) * P(B) = (1/2)*(2/3) = 1/3$ → (II)

From (I), (II): $P(A \cap B) = P(A) * P(B)$ ➔ A, B are Independent

Independent Events

The events $E_1, E_2, \dots, E_n$ are independent if:

$P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_1) * P(E_2) * \dots * P(E_n)$

Example

A pack contains 4 blue, 2 red and 3 black pens. If a pen is drawn at random from the pack, replaced and the process repeated 2 more times, what is the probability of drawing 2 blue pens and 1 black pen in this order?

Solution

P(EBlue) = 4/9

P(EBlack) = 3/9

$$
\begin{aligned}
P(EBlue \cap EBlue \cap EBlack) &= P(EBlue) * P(EBlue) * P(EBlack) \\
&= 4/9 * 4/9 * 3/9 \\
&= 48/729 \\
&= 16/243
\end{aligned}
$$

Bayes theory

$P(A \cap B) = P(A|B)P(B)$ I

$P(B \cap A) = P(B|A)P(A)$ II

From I and II

$P(A|B)P(B) = P(B|A)P(A)$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Example:

What is the patient's probability of having liver disease if they are an alcoholic, knowing that:

10% of patients entering your clinic have liver disease.

Five percent of the clinic's patients are alcoholics.

Bayes theory

among those patients diagnosed with liver disease, 7% are alcoholics.

Solution

A : having liver disease

B : alcoholic patient

$P(A) = 0.1$

$P(B) = 0.05$

$P(B|A) = 0.07$

$$P(A|B) = P(B|A)\, P(A) / P(B)$$
$$= 0.07 * 0.1 / 0.05$$
$$= 0.14$$

Bayes theory

If we have a lot of events, we can use the Total Probability Rule, for P(B), so the general bayes theory is:

If $E_1, E_2, \ldots, E_k$ are K mutually exclusive and exhaustive events and B is any event,

$B = (B \cap E1) \cup (B \cap E2) \cup \ldots \cup (B \cap Ek)$

$P(B) = P(B \cap E1) + P(B \cap E2) + \ldots + P(B \cap Ek)$

$$P(E_1|B) = \frac{P(B|E_1)P(E_1)}{P(B|E_1)P(E_1) + P(B|E_2)P(E_2) + \ldots P(B|E_k)P(E_k)}$$