THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Analysis of Engineering and Scientific Data

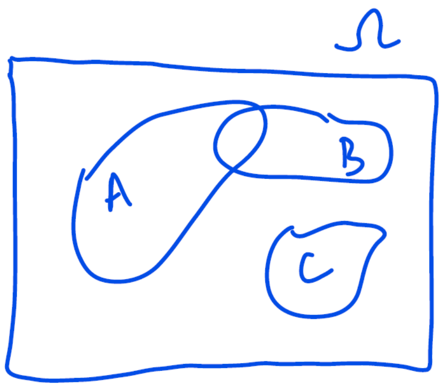
Semester 1 – 2019

Sabrina Streipert s.streipert@uq.edu.au

Lecture 2

1. Sample Space, Outcomes and Events
2. Probability
3. Conditional Probability and Independence

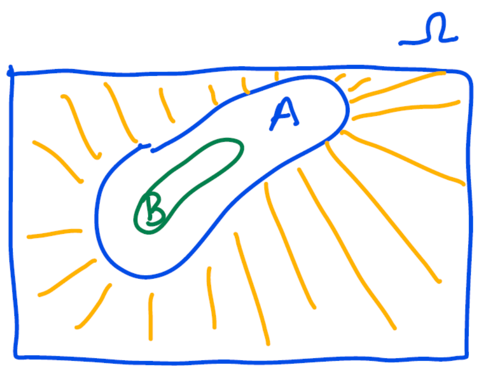
Venn Diagram



A, B, C are events

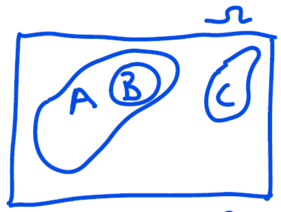
Given sample space Ω and a set $B \subset A$. Highlight A^c in a Venn diagram.

= everything but A = $\Omega \setminus A$



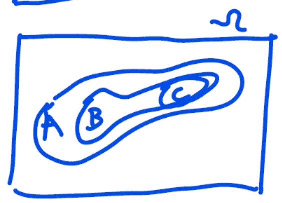
Given sample space Ω and a set $B \subset A$ and C. Highlight $A \cap B \cap C$ in a Venn diagram.

a)



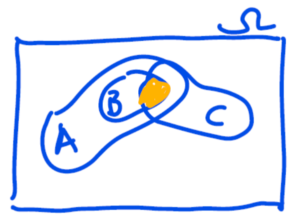
$\Rightarrow A \cap B \cap C = \emptyset$

b)



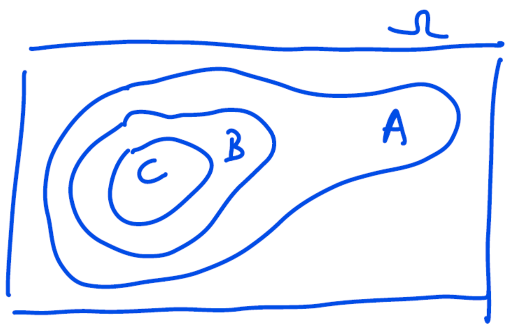
$\Rightarrow A \cap B \cap C = C$

c)



$A \cap B \cap C$

If $C \subset B$ and $B \subset A$, what can we say about the relation between C and A?

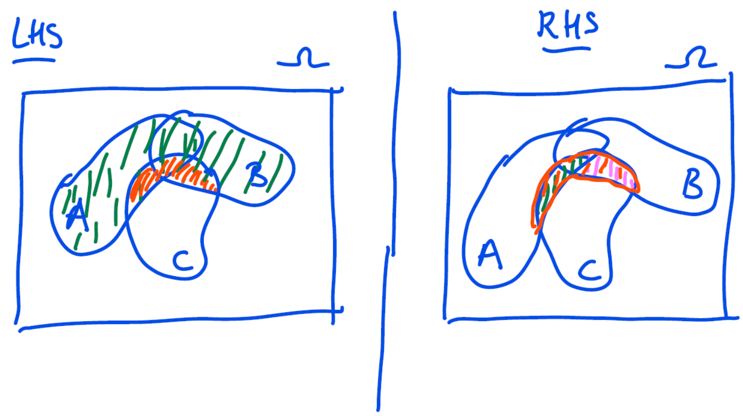


i) $C \subset A$

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

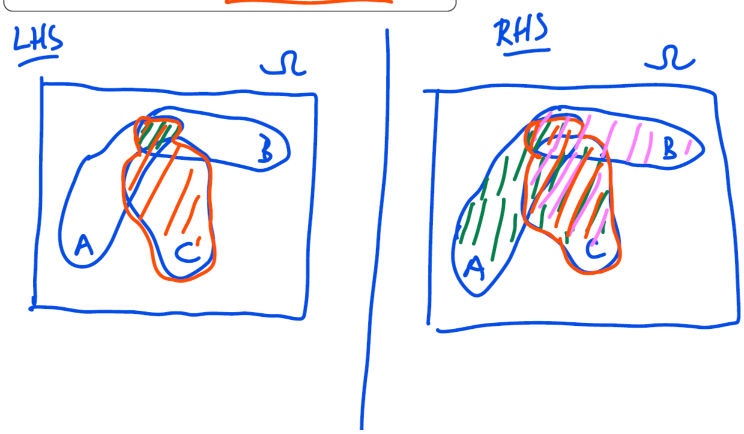
Operation Law - 1

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



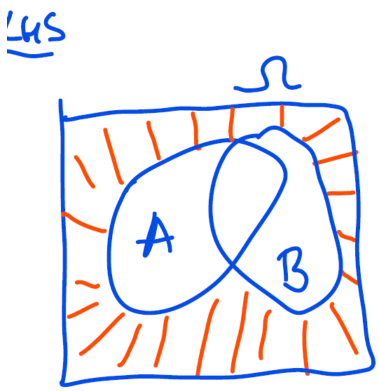
Operation Law - 2

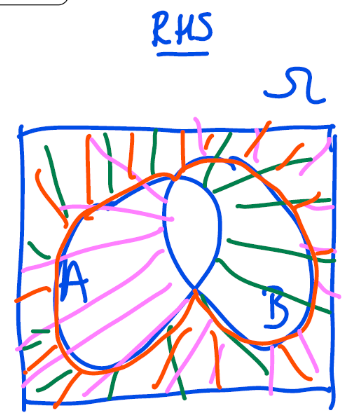
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$



(1st De Morgan's Law)

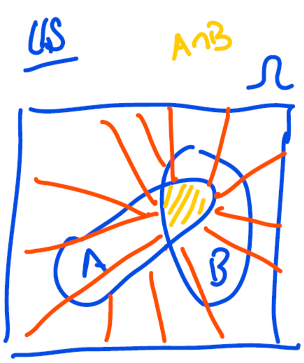
$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

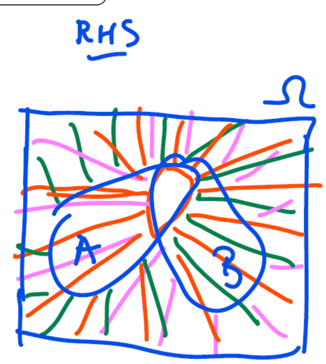




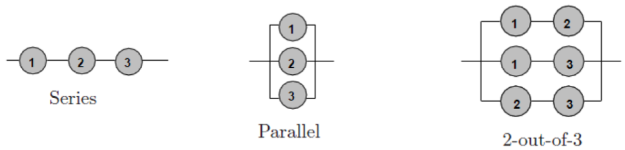
(2nd De Morgan's Law)

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$





Example - Network



Three unreliable systems

Left: Series system works only if all components work.

Center: Parallel system works if at least 1 component works.

Right: 2-out-of-3 system works if at least 2 components work.

Let A_i be the event that the i^{th} component works.

Express the event that "Series system" works using A_i.

$\underline{A_1 \cap A_2 \cap A_3}$

Express the event that "Parallel system" works using A_i.

$\underline{A_1 \cup A_2 \cup A_3}$

Express the event that "2-out-of-3 system" works using A_i.

$\underline{(A_1 \cap A_2) \cup (A_1 \cap A_3) \cup (A_2 \cap A_3)}$

Probability = Likelihood

Probability is a measure of how likely the outcome of an experiment is.
Example:

Sample space Ω is the income of all aerospace engineers in Australia.
Let A be the event that the income of at least 90% of engineers is above 150K a year.

To measure the likelihood of this event A, we ask: *What is the probability/likelihood that at least 90% of engineers have an income of above 150K a year?*

Probability - Formal definition

Definition

Let $\mathcal{F}$ be the set of all subsets of Ω.
A probability $P : \mathcal{F} \to [0, 1]$ such that the following holds:

1. $0 \leq P(A) \leq 1$ for any $A \in \mathcal{F}$
2. $P(\Omega) = 1$
3. If $A_1 \cap A_2 = \emptyset$ then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$
   Generalized:

   If $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

$$P(\cup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P(A_i)$$

Probability - Properties

- $P(\emptyset) = \underline{0}$
- $P(\overline{A}) = \underline{1 - P(A)}$
- If $A \subset B$ then $P(A) \leq P(B)$
- $P(A \cup B) = \underline{P(A) + P(B) - P(A \cap B)}$

Probability - Discrete Sample Space

$\Omega = \{a_1, a_2, \dots, a_n\}$, Event $A = \{a_1, a_4, a_6\}$, then

$$P(A) = P(a_1) + P(a_4) + P(a_6)$$

In general:

$$P(A) = \sum_{i\,:\,a_i \in A} P(a_i)$$

Example:
The probability that a machine produces at least 4 products an hour but less than 10 is
P(machine produces 4 prod) + P(machine produces 5 prod) + ...
... + P(machine produces 9 prod).

Probability

How do we get the values for $P(a_i)$?

- e.g. samples from experiments/available data collection
- **geometric probability**
  If all events are equally likely and Ω is finite, then

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

  If Ω discrete, then $|A|$ = number of non-decomposible elements in set/event A.
  If Ω continuous, then $|A|$ = length of the (continuous) interval representing A.

Example - Phone Number

If each digit from 0 to 9 of a 9-digit phone number are equally likely, what is the probability that:

- the phone number contains nine times the number 9?
  $P = \frac{|A|}{|\Omega|} = \frac{1}{10^9}$

- the phone number contains exactly one 2?
  $P = \frac{|B|}{|\Omega|} = \frac{9 \cdot 9^8}{10^9}$

- the phone number contains at least two 5?
  $P = 1 - P(\text{at most two } 5) = 1 - [P(\text{no } 5) + P(\text{one } 5)]$
  $= 1 - \left[\frac{9^9}{10^9} + \frac{9 \cdot 9^8}{10^9}\right]$

Probability - Continuous Sample Space

geometric probability for continuous sample space

$\Omega = [a, b]$ or any other uncountable set (e.g. any subset of $\mathbb{R}$)
Event $A = (a_1, b_1) \subset [a, b]$ and any value in Ω is **equally likely** , then

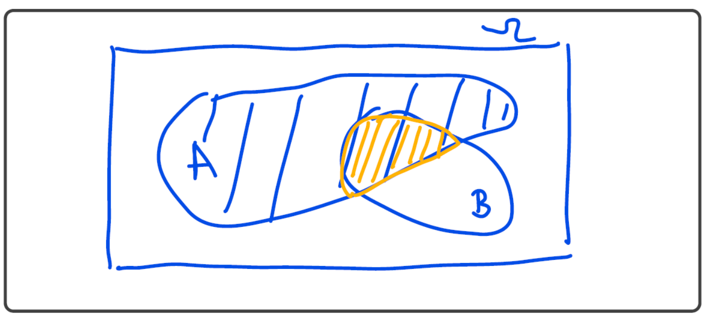
$$P(A) = \frac{|A|}{|\Omega|} = \frac{b_1 - a_1}{b - a}$$

Example:
The probability that a randomly chosen number (not necessarily an integer) between 0 and 100 lies in $(0, 5)$ is

$P(A) = \frac{5-0}{100-0} = \frac{1}{20}$.

Conditional Probability

Suppose we know that event A occurred. How does that affect the probability that B will occur?



⇒ if A happened, then in the shaded area,
so B can only happen in $A \cap B$.

So, the relative change/likelihood that B occurs is

$$P(\text{ B happens, given that A happened}) = \frac{P(A\cap B)}{P(A)}$$

="Probability that B occurs, given that A occurred",denoted by

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

We therefore have

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

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

Example:

Suppose you take two math courses: *Calculus for Engineers* and *Statistics*.
The event A, that the mean grade is 6, is then:
$A = \{(\text{Calc} = 5, \text{Stat} = 7), (\text{Calc} = 6, \text{Stat} = 6), (\text{Calc} = 7, \text{Stat} = 5)\}$.

Let B be the event that you had at least one 7.
Then $A \cap B = \{(\text{Calc} = 5, \text{Stat} = 7), (\text{Calc} = 7, \text{Stat} = 5)\}$.

Given that the mean grade is 6, what is the probability that one grade was a 7?

Considering all cases are equally likely.

$P(B|A) = \underline{P(A\cap B)/P(A) = |A\cap B|/|A| = 2/3}$ $\qquad |\Omega| = \underline{7^2}$

Can we extend that result?

$P(A\cap B\cap C) = \underline{P(C|A,B)\cdot P(B|A)\cdot P(A)}$

Even more general:

$$P(\cap_{i=1}^k A_i) = P(A_k|A_1,\dots,A_{k-1})\cdot P(A_{k-1}|A_{k-2},\dots,A_1)\cdots$$
$$\cdots P(A_2|A_1)P(A_1)$$

Example:

We draw 3 balls from a bowl with 6 white and 4 black balls.

What is the probability that all balls will be black if you do return them after every drawing?

$P = \left(\frac{4}{10}\right)^3$

What is the probability that all balls will be black, if the draws are not returned?

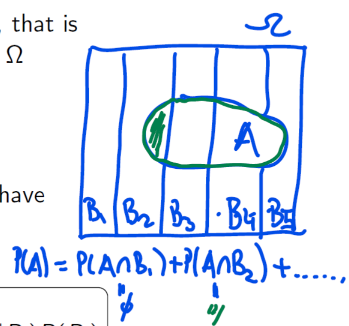
$P = \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8}$

Law of Total Probability

Let $B_1, B_2, \ldots, B_n$ be a **partition** of Ω, that is $B_i \cap B_j = \emptyset$ for all $i \neq j$ **and** $\cup_{i=1}^{n} B_i = \Omega$

Then $P(A) = \sum_{i=1}^{n} P(A \cap B_i)$

and using conditional probabilities, we have



$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Example:

Semiconductor Failures:

P of Failure	Level of Contamination
0.10	High
0.05	Medium
0.01	Low

In a day, 15% of the chips are subjected to high levels of contamination, 25% to medium levels of contamination, and 60% to low levels of contamination.

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

F ="failure", H = "High contamination", M = "Medium contamination", L = "Low contamination". (So partition is the levels of contamination.)

$$\begin{aligned} P(F) = P(F|H)P(H) + P(F|M)P(M) + P(F|L)P(L) \\ = 0.10 \cdot 0.15 + 0.05 \cdot 0.25 + 0.01 \cdot 0.60. \end{aligned}$$

Independence

Two events A and B are **independent** if any of the following holds:

- $P(A \cap B) = \underline{P(A)P(B)}$
- $P(A|B) = \underline{P(A)}$
- $P(B|A) = \underline{P(B)}$

This can be extended:

Events $A_1, A_2, \ldots, A_k$ are independent if

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2)\ldots P(A_k) = \underline{\prod_{i=1}^{k} P(A_i)}$$