INTELLIGENT SYSTEMS LABORATORY

Lesson 2
Basic Probability Theory

Mathematics and Statistics for Data Science

Tapanan Yeophantong
Vincent Mary School of Science and Technology
Assumption University

Content

- Probabilities & probability spaces
- Conditional probabilities & Bayes theorem

Statistical Methods

- Basic Idea: To make inferences about a *population* by studying a relatively small *sample* chosen from it.
- A population is the entire collection of objects or outcomes about which information is sought.
- A sample is a subset of a population, containing the objects or outcomes that are actually observed.

Probability

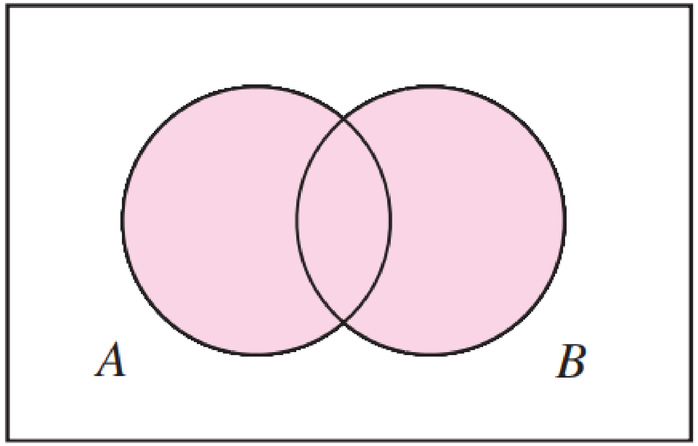
What’s the likelihood that you will get through this?

Preliminary Concepts

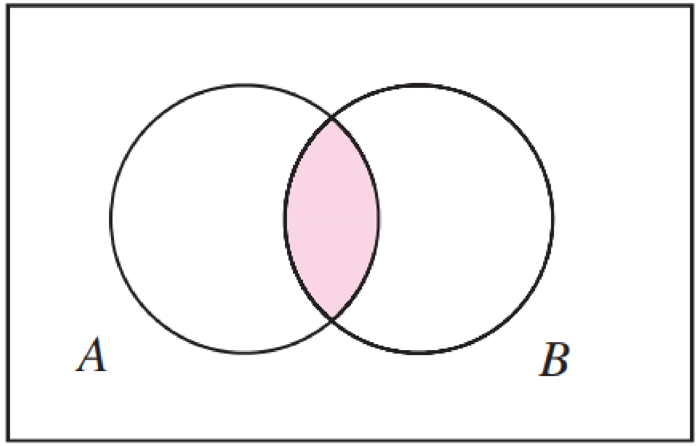
- To understand probability, we start by understanding the terms *sample space* and *event* for an observation.
- The set of all possible outcomes of an experiment is called the sample space for the experiment.
- A subset of a sample space is called an event.

Combining Events

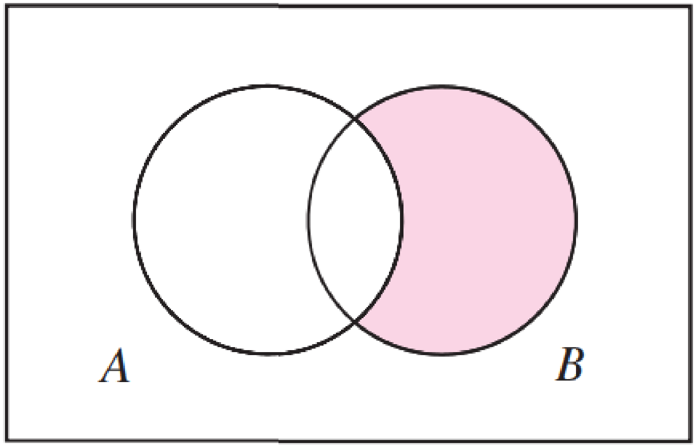
$A \cup B$



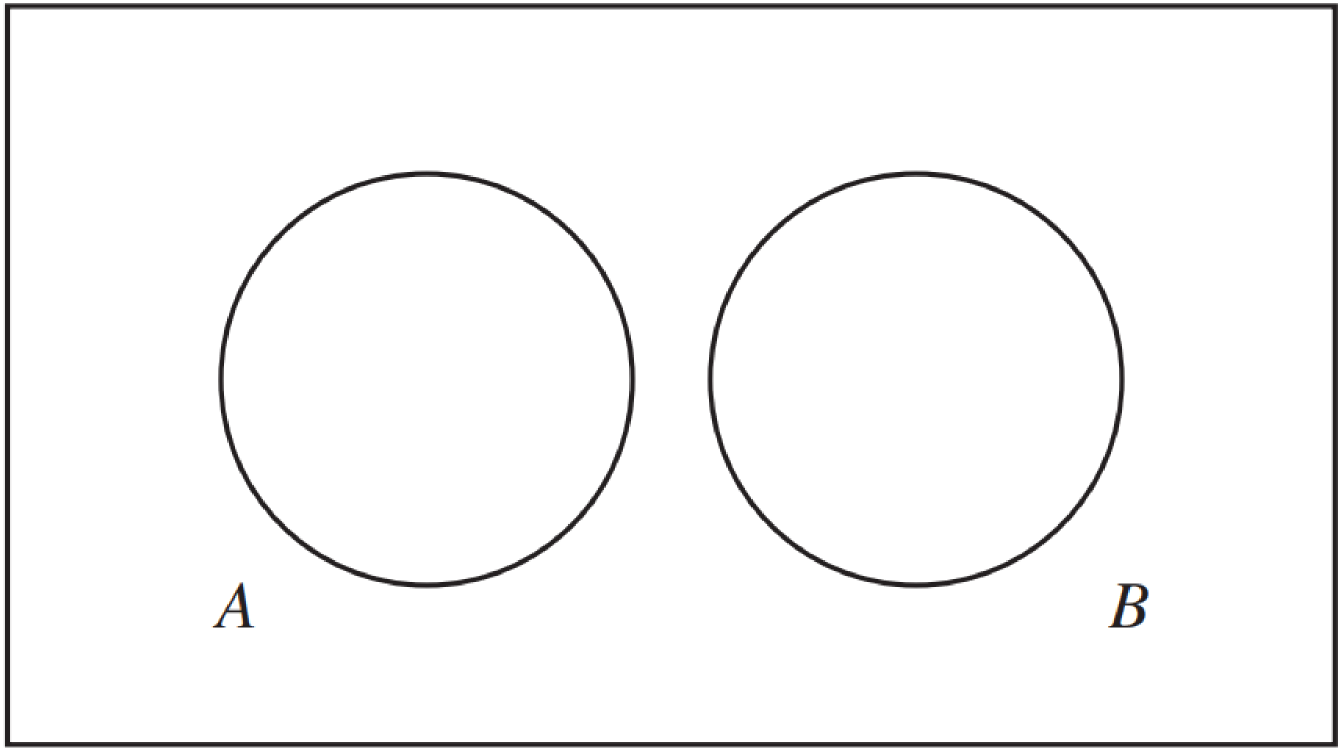
$A \cap B$



$B \cap Ac$



Mutually Exclusive Events



Probabilities

- Given any experiment and any event A:
  - The expression P(A) denotes the probability that the event A occurs.
  - P(A) is the proportion of times that event A would occur in the long run, if the experiment were to be repeated over and over again.
- Axioms of Probability:
  - Let S be a sample space. Then $P(S) = 1$.
  - For any event A, $0 \leq P(A) \leq 1$
  - If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$

Example 1

- From a population of 1,000 rods, a rod is sampled at random.
  - Find P(too short)
  - Find P(too short AND too thick)
  - Find P(too short OR too thick)

	Diameter		
Length	Too Thin	OK	Too Thick
Too Short	10	3	5
OK	38	900	4
Too Long	2	25	13

Example 1 - Solution

- From a population of 1,000 rods, a rod is sampled at random.
  - P(too short) = 18/1000 (0.018)
  - P(too short AND too thick) = 5/1000 (0.005)
  - P(too short OR too thick) = 35/1000 (0.035)

	Diameter		
Length	Too Thin	OK	Too Thick
Too Short	10	3	5
OK	38	900	4
Too Long	2	25	13

Addition Rule

- If A and B are *mutually exclusive* events, then:

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

- More generally, let A and B be *any* events, then:

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

Counting

How many ways can you do things?

Permutation

- A permutation is an ordering of a collection of objects.
- The number of permutations of n objects is n!
- The number of permutations of k objects chosen from a group of n objects is given by:

$$\frac{n!}{(n-k)!}$$

Example 2

- When ordering a certain type of computer, there are 3 choices of hard drive, 4 choices for the amount of memory, 2 choices of video card, and 3 choices of monitor.

- How many ways can a computer be ordered?

Example 2 - Solution

- When ordering a certain type of computer, there are 3 choices of hard drive, 4 choices for the amount of memory, 2 choices of video card, and 3 choices of monitor.

- The number of ways to order a computer is $(3)(4)(2)(3) = 72$.

Combination

- A combination is a selection of distinct group of objects, without regards to order.
- The number of combinations of k objects chosen from a group of n objects is given by:

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

Example 3

- Five lifeguards are available for duty one Saturday afternoon. There are three lifeguard stations.
- In how many ways can three lifeguards be chosen and ordered among the stations?

Example 3 - Solution

- Five lifeguards are available for duty one Saturday afternoon. There are three lifeguard stations.
- In how many ways can three lifeguards be chosen and ordered among the stations?
- Answer: 60

Example 4

- At a certain event, 30 people attend, and 5 will be chosen at random to receive door prizes. The prizes are all the same, so the order in which the people are chosen does not matter.
- How many different groups of five people can be chosen?

Example 4 - Solution

- At a certain event, 30 people attend, and 5 will be chosen at random to receive door prizes. The prizes are all the same, so the order in which the people are chosen does not matter.
- How many different groups of five people can be chosen?

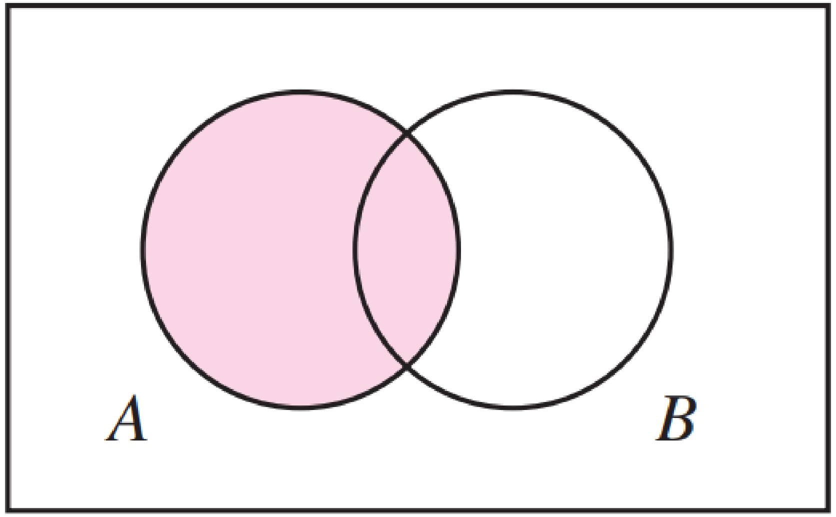
$$\binom{30}{5} = \frac{30!}{5!25!} = \frac{(30)(29)(28)(27)(26)}{(5)(4)(3)(2)(1)} = 142{,}506$$

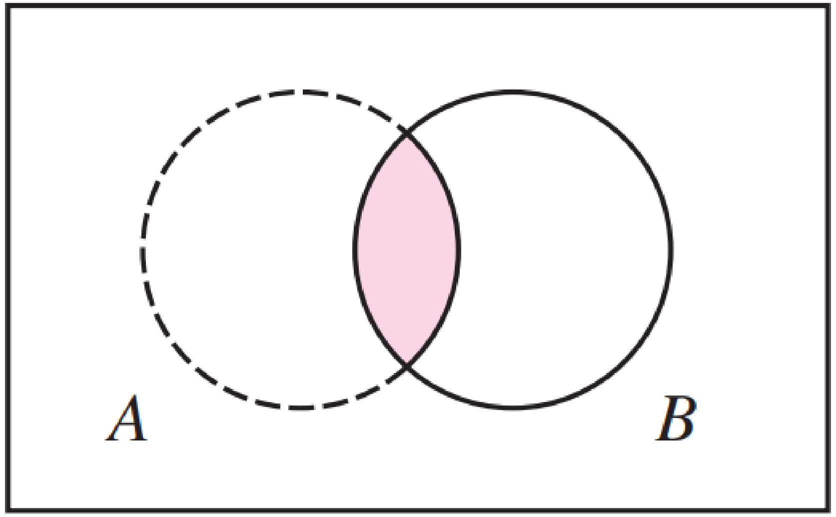
Conditional Probability

- A probability that is based on a part of a sample space.
- Let A and B be events with $P(B) \neq 0$. The conditional probability of A given B, denoted $P(A|B)$, is:

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

Unconditional vs. Conditional





Independence

- Sometimes, knowledge that one event has occurred does not change the probability that another event occurs.
- In this case, the events are said to be *independent*.
- If $P(A) \neq 0$ and $P(B) \neq 0$, then A and B are independent if:

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

and equally, $P(A|B) = P(A)$

Bayes Theorem

- Bayes Theorem provides a formula that allows us to calculate one of the conditional probabilities if we know the other one.
- By definition of conditional probability:

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

- We can substitute P(B|A)P(A) for P(A ∩ B):

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

Example 5

- Suppose the proportion of people who have COVID-19 is 0.005. A test is available to diagnose the disease. If a person has the disease, the probability that the test will produce a positive signal is 0.99. If a person does not have the disease, the probability that the test will produce a positive signal is 0.01.
- If a person tests positive, what is the probability that the person actually has the disease?

Example 5 - Solution

- Let D represent the event that the person actually has the disease, and let + represent the event that the test gives a positive signal.
- We want to find P(D|+). We know:

$$P(D) = 0.005 \qquad P(+|D) = 0.99 \qquad P(+|D^c) = 0.01$$

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} \\ &= \frac{(0.99)(0.005)}{(0.99)(0.005) + (0.01)(0.995)} \\ &= 0.332 \end{aligned}$$