Basics of Probability

Angshuman Paul

Assistant Professor

Department of Computer Science & Engineering

|| त्वं ज्ञानमयो विज्ञानमयोऽसि ||

Uncertainty

- My statement: ***Real world is full of uncertainties***
  - Is it?
  - What about a coin tossing (is it deterministic)
  - Can we deterministically model it?

Uncertainty

- My statement: ***Real world is full of uncertainties***
  - Is it?
  - What about a coin tossing (is it deterministic)
  - Can we deterministically model it?
    - If we know all the physics
    - But it is very hard to model everything

Uncertainty

- My statement: ***Real world is full of uncertainties***
  - Is it?
  - What about a coin tossing (is it deterministic)
  - Can we deterministically model it?
    - If we know all the physics
    - But it is very hard to model everything

- At the quantum label, the world is probabilistic
- So, it is really a philosophical question whether the world is deterministic or probabilistic

Uncertainty

- However, even in the coin tossing example, we usually don’t model all the physics
  - Difficulty in modeling
- In practice, we often don’t model the exact system
  - We use some abstractions
    - That results in the probabilistic nature of our experiments
  - And use probability to model the system

Probability

Experiments

- Deterministic
  - Outcome is known
    - If you release an object from the roof of your house, it will always move in the downward direction (and not in the upward direction)
- Random
  - Outcome is not known beforehand
    - Rolling a dice

Sample Space

- The set of possible outcomes
- Rolling a dice
  - Sample space $S = \{1, 2, 3, 4, 5, 6\}$
  - Each element of sample space: sample point

- Sample space can be
  - Finite (above example)
  - Infinite (Amount of rainfall at Jodhpur in July)

Event

- Any subset of the sample space
- Example: In the context of rolling a dice
  - The outcome is a number < 3
  - Event $A = \{1, 2\}$

Random Variables

- In logic, we have symbols

- In probability, we have random variables (rv)
  - A numerical description of the outcomes of a random experiment
  - A function that assigns numerical values (real or Boolean) to each sample point
  - Usually indicated using capital letters (e.g., X)
    - Discrete: takes only a countable number of discrete values
      - Sample space for weather condition: $\{sunny, rainy, cloudy\}$
    - Continuous: takes uncountably infinite number of possible values
      - Sample space for temperature at Jodhpur: $[6.7^o, 46.2^o]$

Probability of an Event

- Consider an event A
- Probability of event A
  - $P(A) = \frac{Number\ of\ elements\ in\ set\ A}{Number\ of\ elements\ in\ the\ sample\ space\ S}$
  - $P(A) = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ outcomes}$
- Example: In the context of rolling a dice
  - Event A: The outcome is an odd number
  - Event $A = \{\ldots\}$

Probability of an Event

- Probability of event A
  - $P(A) = \frac{Number\ of\ elements\ in\ set\ A}{Number\ of\ elements\ in\ the\ sample\ space\ S}$
  - $P(A) = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ outcomes}$
- Example: In the context of rolling a dice
  - Event A: The outcome is an odd number
  - Event $A = \{1, 3, 5\}$

Probability of an Event

- Probability of event A
  - $P(A) = \frac{Number\ of\ elements\ in\ set\ A}{Number\ of\ elements\ in\ the\ sample\ space\ S}$
  - $P(A) = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ outcomes}$
- Example: In the context of rolling a dice
  - Event A: The outcome is an odd number
  - Event $A = \{1, 3, 5\}$
  - $P(A) = \frac{3}{6} = 0.5$

Probability of an Event

- Probability of event A
  - $P(A) = \frac{Number\ of\ elements\ in\ set\ A}{Number\ of\ elements\ in\ the\ sample\ space\ S}$
  - $P(A) = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ outcomes}$
- Example: In the context of rolling a dice
  - Event A: The outcome is an odd number
  - Event $A = \{1, 3, 5\}$
  - $P(A) = \frac{3}{6} = 0.5$

What is the problem with this definition?

Probability of an Event

- In an experiment with **finite sample space and equally likely outcomes**, Probability of event A
  - $P(A) = \frac{Number\ of\ elements\ in\ set\ A}{Number\ of\ elements\ in\ the\ sample\ space\ S}$
  - $P(A) = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ outcomes}$
- Example: In the context of rolling a dice
  - Event A: The outcome is an odd number
  - Event $A = \{1, 3, 5\}$
  - $P(A) = \frac{3}{6} = 0.5$

Probability of an Event

- In an experiment with **finite sample space and equally likely outcomes**, Probability of event A
  - $P(A) = \frac{Number\ of\ elements\ in\ set\ A}{Number\ of\ elements\ in\ the\ sample\ space\ S}$
  - $P(A) = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ outcomes}$
- Example: In the context of rolling a dice
  - Event A: The outcome is an odd number
  - Event $A = \{1, 3, 5\}$
  - $P(A) = \frac{3}{6} = 0.5$

What if this condition does not hold?

Frequentist Approach of Probability

- Suppose we do an experiment n number of times
- Out of these, event A occurs $n(A)$ number of times

- Relative frequency of A is $f_r(A) = \frac{n(A)}{n}$
- Probability of A is $P(A) =$

Frequentist Approach of Probability

- Suppose we do an experiment n number of times
- Out of these, event A occurs $n(A)$ number of times

- Relative frequency of A is $f_r(A) = \frac{n(A)}{n}$
- Probability of A is $P(A) = \lim_{n \to \infty} f_r(A) = \lim_{n \to \infty} \frac{n(A)}{n}$

Probability

- A probability measure or probability function $P(\cdot)$ assigns a probability to an event
  - $P(A)$ is the chance that event A occurs

Probability: Properties

- $P(A)$ must be a positive number between 0 and 1 inclusive $(0 \leq P(A) \leq 1)$

- If Φ is the null event (indicating no outcome from an experiment), $P(\Phi) = 0$
  - It is impossible that an experiment has no outcome

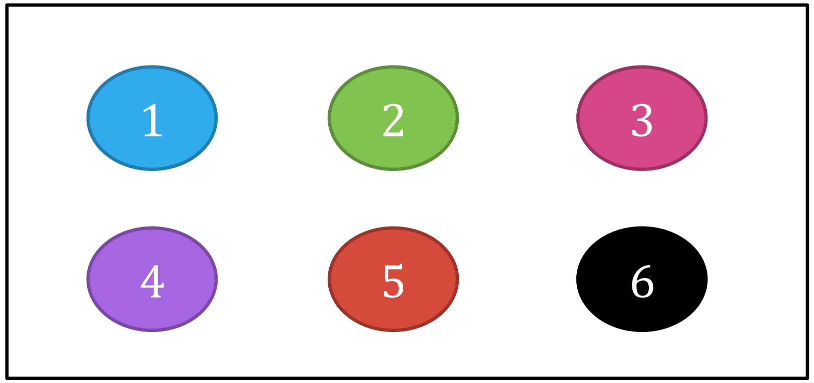
- If S is the sample space, $P(S) = 1$
  - Probability that something happens is always 1

- If $A_1, A_2, \ldots, A_n$ are a countable sequence of disjoint events, then
  - $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

Probability: Properties

- Disjoint/ Mutually exclusive: Two events are disjoint or mutually exclusive, if they can't occur simultaneously

- $P(2 \cup 4 \cup 5) = P(2) + P(4) + \ P(5)$
  - We say it like this: ***If we roll a dice, the probability of getting a 2 or a 4 or a 5 is the sum of the probability of getting a 2, probability of getting a 4, and probability of getting a 5***

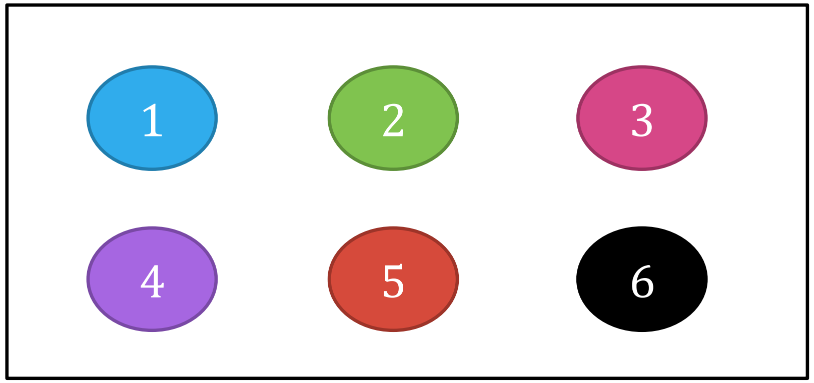


Probability: Properties

- If $A_1, A_2, \dots, A_n$ are a countable sequence of disjoint events, then
  - $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$

- Disjoint/ Mutually exclusive: Two events are disjoint or mutually exclusive, if they can't occur simultaneously

- Example: In case of rolling of dice

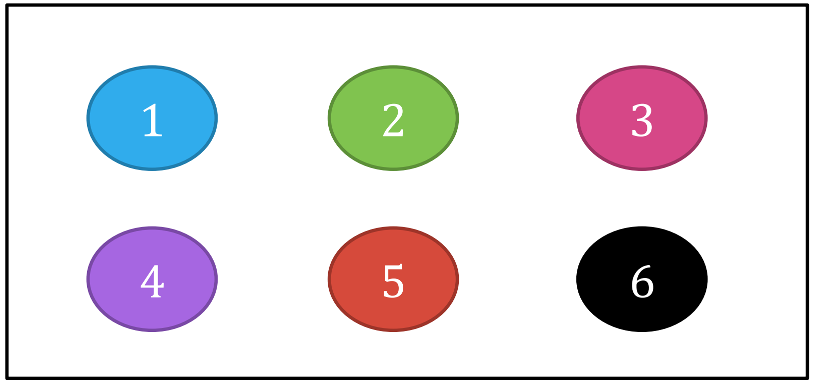


Probability: Properties

- If $A_1, A_2, \ldots, A_n$ are a countable sequence of disjoint events, then
  - $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

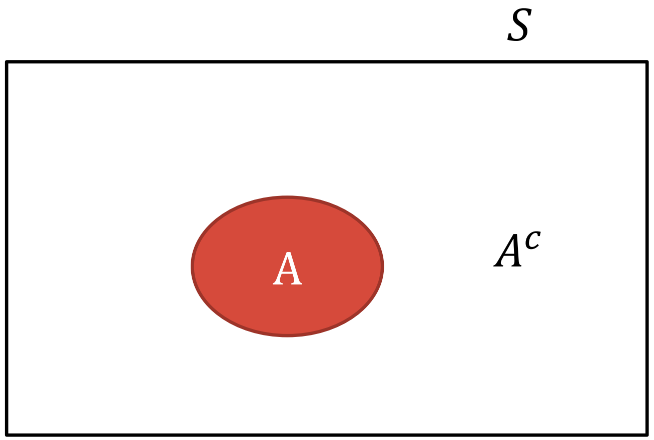
- Disjoint/ Mutually exclusive: Two events are disjoint or mutually exclusive, if they can't occur simultaneously

- Example: Rolling of dice



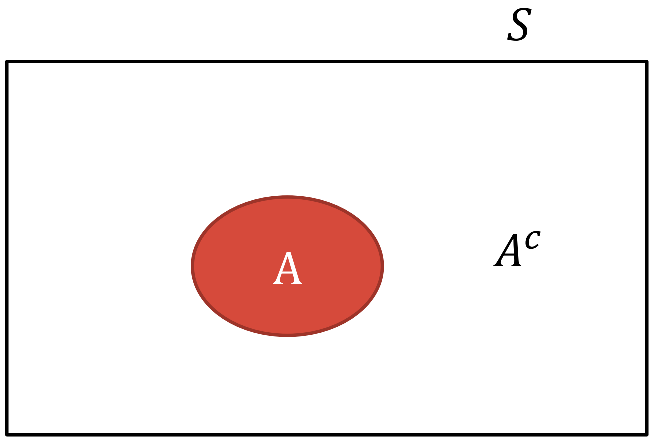
Venn Diagram

Probability: Properties



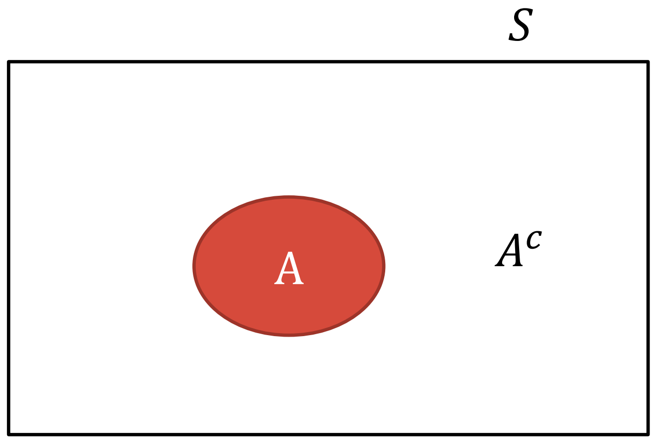
- $S =$

Probability: Properties



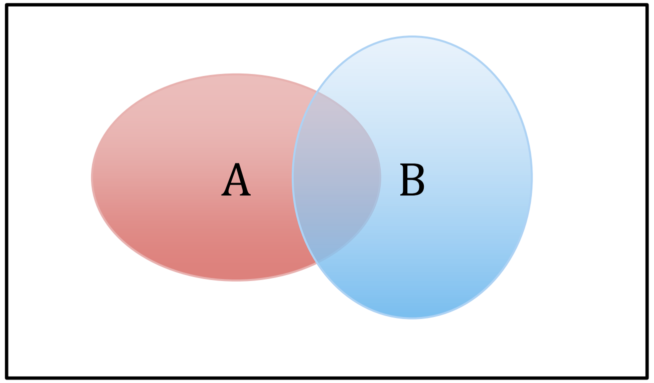
- $S = A \cup A^c$
- $P(S) = 1 = P(A \cup A^c) =$

Probability: Properties

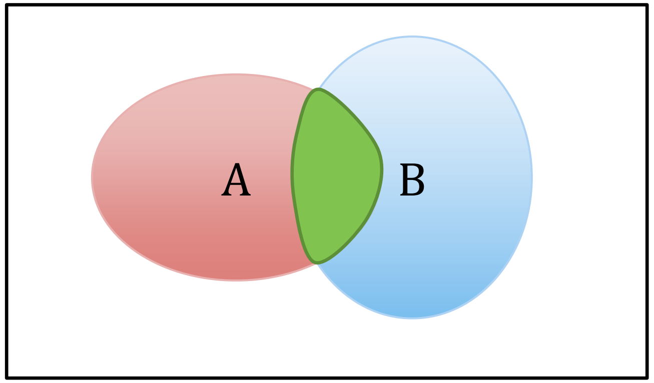


- $S = A \cup A^c$
- $P(S) = 1 = P(A \cup A^c) = P(A) + P(A^c)$
- $P(A) = 1 - P(A^c)$

Probability: Addition Rule



Probability: Addition Rule



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

Revisiting Random Variables

- X is a discrete rv with k possible values $x_1, x_2, \ldots, x_k$
- $P(X = x_j) = P_j$
- Then $\sum_{i=1}^{k} P_i = 1$

- Probability distribution: A list containing $P_1, P_2, \ldots, P_k$

- Often represented as histogram

Revisiting Random Variables

- If X is a continuous rv
  - The probabilities of the value of X are represented by a curve $f(v)$
    - $f(v) \geq 0$ at all points
    - Area under the curve is 1
- $P(a < X \leq b) = \int_a^b f(v)dv$
- $\int_{-\infty}^{\infty} f(v)dv = 1$

Conditional Probability

- In a box, there are 40 Samsung phones and 20 MI phones
- Out of these, 10 Samsung phones and 2 MI phones are not working.
- You pick up a phone and find that the phone is not working.

- What is the probability that the phone you picked up is a Samsung phone?

Conditional Probability

- In a box, there are 40 Samsung phones (SP) and 20 MI phones (MIP)
- Out of these, 10 Samsung phones and 2 MI phones are not working (NW).
- You pick up a phone and find that the phone is not working.

- What is the probability that the phone you picked up is a Samsung phone?

- Find $P(SP|NW)$

Conditional Probability

- In a box, there are 40 Samsung phones (SP) and 20 MI phones (MIP)
- Out of these, 10 Samsung phones and 2 MI phones are not working (NW).
- You pick up a phone and find that the phone is not working.
- What is the probability that the phone you picked up is a Samsung phone?
- Find $P(SP|NW)$

- $P(SP|\text{NW}) = \dfrac{\#\, Samsung\ phones\ that\ are\ not\ working}{\#\, Phones\ that\ are\ not\ working}$

$$= \frac{10}{12}$$

$$= \frac{\frac{10}{40} \times \frac{40}{60}}{\frac{12}{60}}$$

- $\frac{10}{40}$: Given an SP, probability that it is NW ($P(NW|SP)$)
- $\frac{40}{60}$: Probability of SP in the box $P(SP)$
- $\frac{12}{60}$: Probability of finding a NW phone in the box $P(NW)$

Conditional Probability

- In a box, there are 40 Samsung phones (SP) and 20 MI phones (MIP)
- Out of these, 10 Samsung phones and 2 MI phones are not working (NW).
- You pick up a phone and find that the phone is not working.

- What is the probability that the phone you picked up is a Samsung phone?

- Find $P(SP|NW)$

- $P(SP|\mathrm{NW}) = \dfrac{\#\, Samsung\ phones\ that\ are\ not\ working}{\#\, Phones\ that\ are\ not\ working}$

$$= \frac{10}{12}$$

$$= \frac{\frac{10}{40} \times \frac{40}{60}}{\frac{12}{60}}$$

- $\frac{10}{40}$: Given an SP, probability that it is NW ($P(NW|SP)$)
- $\frac{40}{60}$: Probability of SP in the box $P(SP)$
- $\frac{12}{60}$: Probability of finding a NW phone in the box $P(NW)$

- $P(SP|\mathrm{NW}) = \dfrac{P(NW|SP)\, P(SP)}{P(NW)}$

Conditional Probability

- In a box, there are 40 Samsung phones (SP) and 20 MI phones (MIP)
- Out of these, 10 Samsung phones and 2 MI phones are not working (NW).
- You pick up a phone and find that the phone is not working.

- What is the probability that the phone you picked up is a Samsung phone?

- Find $P(SP|NW)$

- $P(SP|\mathrm{NW}) = \dfrac{\#\ Samsung\ phones\ that\ are\ not\ working}{\#\ Phones\ that\ are\ not\ working}$
- $P(SP|\mathrm{NW}) = \dfrac{\#\ Samsung\ phones\ that\ are\ not\ working/60}{\#\ Phones\ that\ are\ not\ working/60}$
- $P(SP|\mathrm{NW}) = \dfrac{P(SP \cap NW)}{P(NW)}$

- We have already seen that

$$P(SP|\mathrm{NW}) = \frac{P(NW|SP)\ P(SP)}{P(NW)}$$

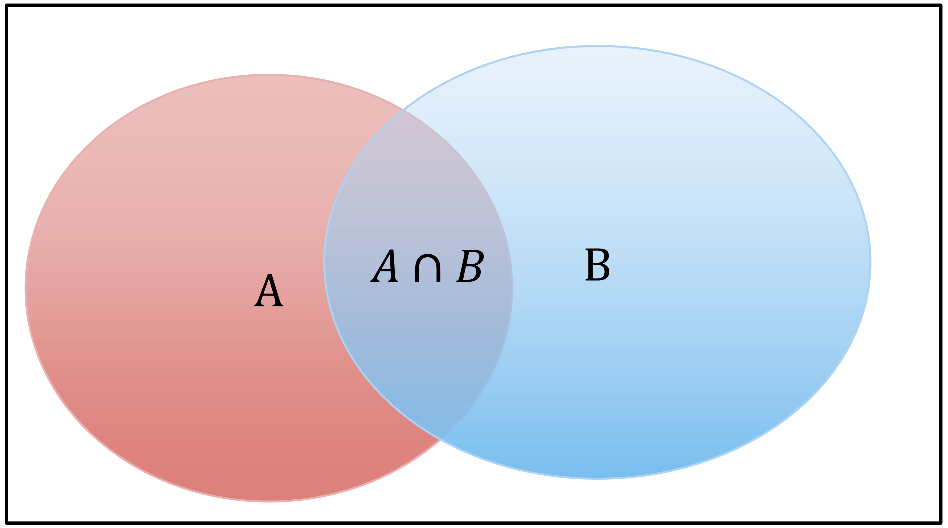
Conditional Probability

- $P(SP|\mathrm{NW}) = \frac{P(SP \cap NW)}{P(NW)} = \frac{P(NW|SP)\, P(SP)}{P(NW)}$

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

Conditional Probability

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

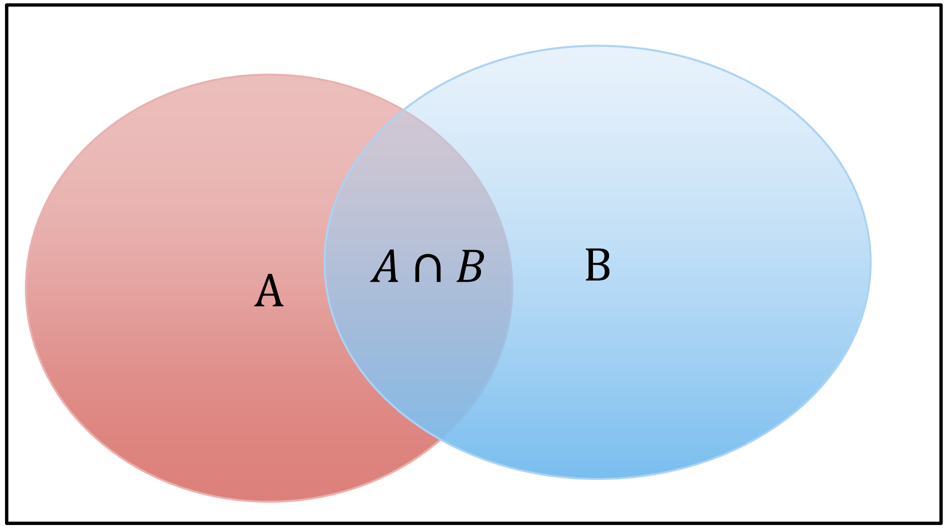


Conditional Probability

Prior Probability

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

Indicates my belief about the occurrence of A

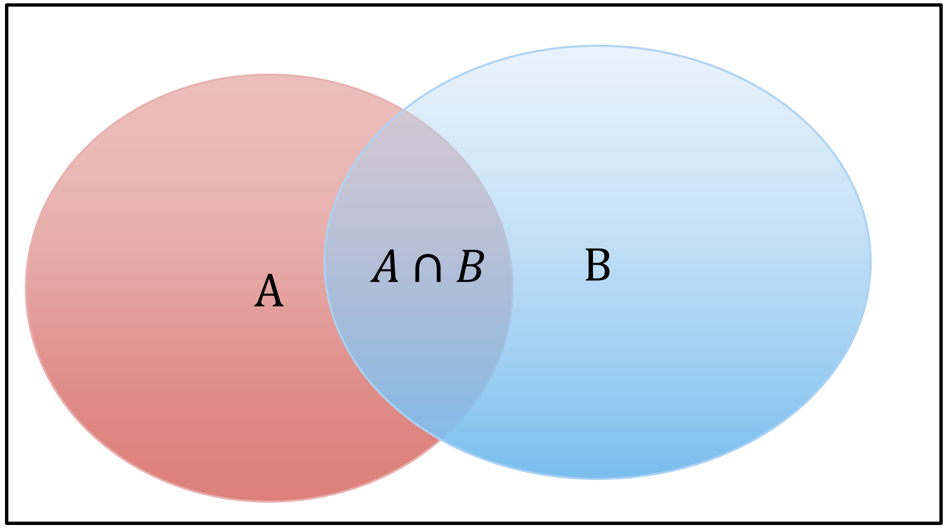


Conditional Probability

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

Likelihood

Indicates the chance of B to occur given that A has occurred

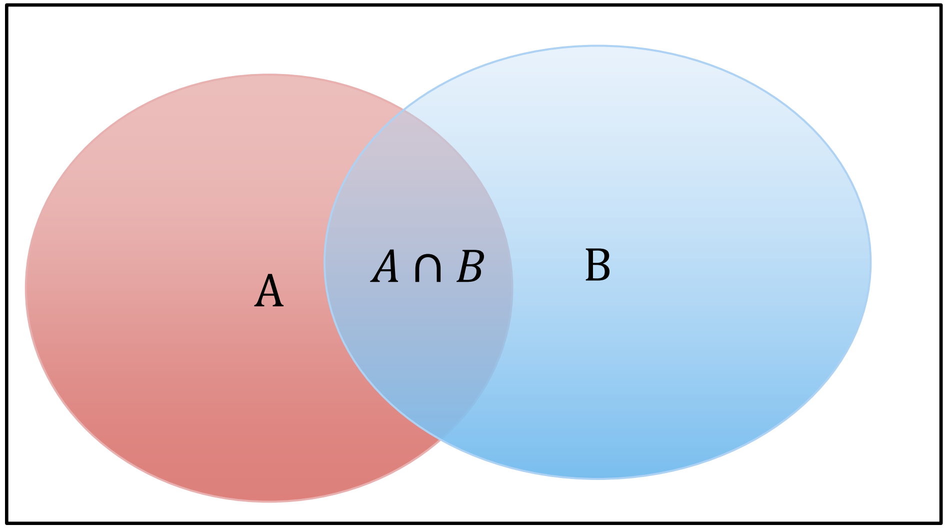


Conditional Probability

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

Evidence

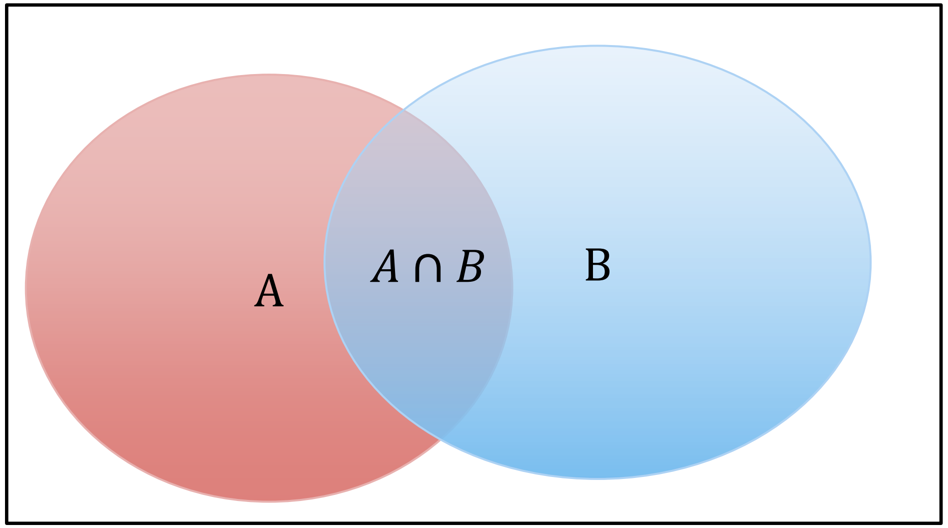
Probability that B occurs



Conditional Probability

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

Posterior/
Conditional
probability

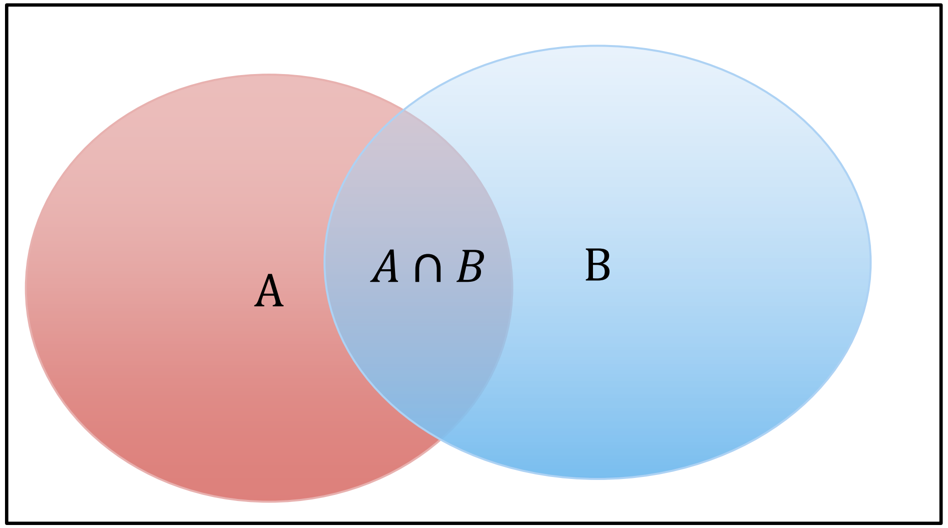


Conditioned on the evidence that we have seen

Conditional Probability

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

Bayes' Theorem



Conditional Probability

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

Bayes' Theorem

$$P(Cause|Effect) = \frac{P(Effect|Cause)\,P(Cause)}{P(Effect)}$$

Conditional Probability

- Consider a system with causes c1, and c2, and effects e1, e2 and e3

- Suppose we want to find the probabilities of different causes given effect e2

$$P(Cause|Effect) = \frac{P(Effect|Cause)\,P(Cause)}{P(Effect)}$$

Conditional Probability

- Consider a system with causes c1, and c2, and effects e1, e2 and e3
- Suppose we want to find the probabilities of different causes given effect e1

$$P(Cause|Effect) = \frac{P(Effect|Cause)\,P(Cause)}{P(Effect)}$$

$$P(c1|e1) = \frac{P(e1|c1)\,P(c1)}{P(e1)}$$

$$P(c2|e1) = \frac{P(e1|c2)\,P(c2)}{P(e1)}$$

Conditional Probability

- Consider a system with causes c1, and c2, and effects e1, e2 and e3
- Suppose we want to find the probabilities of different causes given effect e1
- Since e1 may be caused either by c1 or by c2
  - $\boldsymbol{P(c1|e1) + P(c2|e1) = 1}$

$$\boldsymbol{P(c1|e1) = \frac{P(e1|c1)\,P(c1)}{P(e1)}} \qquad \boldsymbol{P(c2|e1) = \frac{P(e1|c2)\,P(c2)}{P(e1)}}$$

Conditional Probability

- Since e1 may be caused either by c1 or by c2
  - $\boldsymbol{P(c1|e1) + P(c2|e1) = 1}$ (1)

- For both $\boldsymbol{P(c1|e1)}$ and $\boldsymbol{P(c2|e1)}$, the denominator is same
  - Consider $\boldsymbol{\frac{1}{P(e1)} = \alpha}$
  - Then, even if we don't calculate $\boldsymbol{P(e1)}$ explicitly, we can find $\boldsymbol{P(c1|e1)}$ and $\boldsymbol{P(c2|e1)}$ using (1)

$$\boldsymbol{P(c1|e1) = \frac{P(e1|c1)\,P(c1)}{P(e1)}}$$

$$\boldsymbol{P(c2|e1) = \frac{P(e1|c2)\,P(c2)}{P(e1)}}$$

Conditional Probability

$$P(c1|e1) = \alpha P(e1|c1)P(c1)$$

$$P(c2|e1) = \alpha P(e1|c2)P(c2)$$

$$P(c1|e1) \propto P(e1|c1)P(c1)$$

$$P(c2|e1) \propto P(e1|c2)P(c2)$$

- So
  - $Posterior \propto (likelihood \times prior)$

Independent Events

- Two events A and B are said to be independent if
  - $\boldsymbol{P(A|B) = P(A)}$ (1)

- We already have
  - $\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)\,P(A)}{P(B)}}$ (2)

- From (1) and (2), we get
  - $\boldsymbol{P(B|A) = P(B)}$
  - $\boldsymbol{P(A \cap B) = P(A)P(B)}$

Joint Probability Distribution

- Consider two random variables
- X corresponding to weather $\{sunny, rainy, cloudy\}$
  - $P(X) = \{0.6, 0.1, 0.3\}$
- Y corresponding to power cut $\{power\ cut, no\ power\ cut\}$
  - $P(Y) = \{0.15, 0.85\}$
- A joint probability distribution of $X\ and\ Y$
  - Probability distribution on all possible pairs of outputs

Joint Probability Distribution

- X corresponding to weather $\{sunny, rainy, cloudy\}$
  - $P(X) = \{0.6, 0.1, 0.3\}$
- Y corresponding to power cut $\{power\ cut, no\ power\ cut\}$
  - $P(Y) = \{0.15, 0.85\}$
- A joint probability distribution of $X\ and\ Y$
  - Probability distribution on all possible pairs of outputs
- A 3×2 matrix of values

Joint Probability Distribution

- X corresponding to weather $\{sunny, rainy, cloudy\}$
  - $P(X) = \{0.6, 0.1, 0.3\}$
- Y corresponding to power cut $\{power\ cut, no\ power\ cut\}$
  - $P(Y) = \{0.15, 0.85\}$
- A joint probability distribution of $X\ and\ Y$
  - Probability distribution on all possible pairs of outputs
- A 3×2 matrix of values

	Power cut	**No power cut**
Sunny	0.01	0.4
Rainy	0.2	0.1
Cloudy	0.09	0.2

Joint Probability Distribution

- Sample space corresponding to X is $S_X = \{s, r, c\}$
- Sample space corresponding to Y is $S_Y = \{pc, npc\}$
- Sample space corresponding to the joint distribution is

	Power cut (pc)	**No power cut (npc)**
Sunny (s)	0.01 $P(s \cap pc)$	0.4 $P(s \cap npc)$
Rainy (r)	0.2 $P(r \cap pc)$	0.1 $P(r \cap npc)$
Cloudy (c)	0.09 $P(c \cap pc)$	0.2 $P(r \cap npc)$

Joint Probability Distribution

- Sample space corresponding to X is $S_X = \{s, r, c\}$
- Sample space corresponding to Y is $S_Y = \{pc, npc\}$
- Sample space corresponding to the joint distribution is

$$S_J = \{(s, pc), (s, npc), (r, pc), (r, npc), (c, pc), (c, npc)\}$$

	Power cut (pc)	**No power cut (npc)**
Sunny (s)	0.01 $P(s \cap pc)$	0.4 $P(s \cap npc)$
Rainy (r)	0.2 $P(r \cap pc)$	0.1 $P(r \cap npc)$
Cloudy (c)	0.09 $P(c \cap pc)$	0.2 $P(r \cap npc)$

Chain Rule

- If $A_1, A_2, \dots, A_n$ are n events, then
  - $P(A_n \cap A_{n-1} \cap \dots \cap A_1) = P(A_n | A_{n-1} \cap \dots \cap A_1)\, P(A_{n-1} \cap \dots \cap A_1)$ (1)
- Similarly,
  - $P(A_{n-1} \cap A_{n-2} \cap \dots \cap A_1) = P(A_{n-1} | A_{n-2} \cap \dots \cap A_1)\, P(A_{n-2} \cap \dots \cap A_1)$ (2)
- Extending this for the subsequent events and putting in (1), we get,
  - $P(A_n \cap A_{n-1} \cap \dots \cap A_1)$

    $= P\,(A_n | A_{n-1} \cap \dots \cap A_1)\, P(A_{n-1} | A_{n-2} \cap \dots \cap A_1) P(A_{n-2} | A_{n-3} \cap \dots \cap A_1) \dots P(A_1)$

Chain Rule

- If A_1, A_2, A_3 are 3 events, then we use
  - $P(A_n \cap A_{n-1} \cap \cdots \cap A_1)$

    $= P\ (A_n | A_{n-1} \cap \cdots \cap A_1)\ P(A_{n-1} | A_{n-2} \cap \cdots \cap A_1) P(A_{n-2} | A_{n-3} \cap \cdots \cap A_1) \ldots P(A_1)$

- We get
  - $P(A_4 \cap A_3 \cap A_2 \cap A_1)$

    $= P\ (A_4 | A_3 \cap A_2 \cap A_1)\ P(A_3 | A_2 \cap A_1) P(A_2 | A_1) P(A_1)$

Inference by Enumeration

	fever		**¬ fever**	
	cough	**¬ cough**	**cough**	**¬ cough**
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

Inference by Enumeration

	fever		¬ fever	
	cough	**¬ cough**	**cough**	**¬ cough**
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- $P(fever) =$

Inference by Enumeration

	fever		¬ fever	
	cough	**¬ cough**	**cough**	**¬ cough**
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- For any proposition, add all the boxes where the proposition is true

- $P(fever) = 0.21 + 0.10 + 0.11 + 0.07 = 0.49$

Inference by Enumeration

	fever		¬ fever	
	cough	**¬ cough**	**cough**	**¬ cough**
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- For any proposition, add all the boxes where the proposition is true

- $P(fever \vee covid) =$

Inference by Enumeration

	fever		¬ fever	
	cough	¬ cough	cough	¬ cough
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- For any proposition, add all the boxes where the proposition is true

- $P(fever \vee covid) = 0.21 + 0.10 + 0.11 + 0.07 + 0.11 + 0.08 = 0.68$

Inference by Enumeration

	fever		**¬ fever**	
	cough	**¬ cough**	**cough**	**¬ cough**
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- For any proposition, add all the boxes where the proposition is true
- $P(\neg covid | \neg fever) =$

Inference by Enumeration

	fever		¬ fever	
	cough	**¬ cough**	**cough**	**¬ cough**
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- For any proposition, add all the boxes where the proposition is true

- $P(\neg covid | \neg fever) = \frac{P(\neg covid \cap \neg fever)}{P(\neg fever)} = \frac{0.09+0.23}{P(\neg fever)}$

Inference by Enumeration

	fever		¬ fever	
	cough	**¬ cough**	**cough**	**¬ cough**
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- For any proposition, add all the boxes where the proposition is true
- $P(\neg covid|\neg fever) = \frac{P(\neg covid \cap \neg fever)}{P(\neg fever)} = \frac{0.09+0.23}{0.11+0.08+0.09+0.23} \approx 0.627$

Inference by Enumeration

	fever		¬ fever	
	cough	**¬ cough**	**cough**	**¬ cough**
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- If we have the complete joint distribution, I can answer any related queries
- But, what is the problem with this approach?

Inference by Enumeration

	fever		¬ fever	
	cough	¬ cough	cough	¬ cough
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- But, what is the problem with this approach?
  - For a system with many causes and effects, we have to maintain a large set of values and operate on those

Independent Events

- Two events A and B are said to be independent if
  - $\boldsymbol{P(A|B) = P(A)}$ (1)
- We already have
  - $\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)\,P(A)}{P(B)}}$ (2)
- From (1) and (2), we get
  - $\boldsymbol{P(B|A) = P(B)}$
  - $\boldsymbol{P(A \cap B) = P(A)P(B)}$

Independent Events

- Suppose I want to deal with Fever, Cough, Covid, and Internet Speed (I_Sp)
- We intuitively know that internet speed does not depend on the other three
- So, if we want to find out the joint distribution
  - $P(Fever, Cough, Covid, I\_Sp)$
- We can write

$$P(Fever, Cough, Covid, I\_Sp) = P(Fever, Cough, Covid)\ P(I\_Sp)$$

Independent Events

- Suppose the sample space for Internet Speed (I_Sp) is

$$\{slow, medium, fast, very\ fast\}$$

- So, if we want to find out the joint distribution
  - $P(Fever, Cough, Covid, I\_Sp)$

		fever		¬ fever	
		cough	**¬ cough**	**cough**	**¬ cough**
covid	**slow**	a1	a2	a3	a4
	medium	a5	a6	a7	a8
	fast	a9	a10	a11	a12
	very fast	a13	a14	a15	a16
¬ covid	**slow**	a17	a18	a19	a20
	medium	a21	a22	a23	a24
	fast	a25	a26	a27	a28
	very fast	a29	a30	a31	a32

Independent Events

		fever		¬ fever	
		cough	**¬ cough**	**cough**	**¬ cough**
covid	**slow**	a1	a2	a3	a4
	medium	a5	a6	a7	a8
	fast	a9	a10	a11	a12
	very fast	a13	a14	a15	a16
¬ covid	**slow**	a17	a18	a19	a20
	medium	a21	a22	a23	a24
	fast	a25	a26	a27	a28
	very fast	a29	a30	a31	a32

- How many variables do I need to store this table?

Independent Events

		fever		¬ fever	
		cough	**¬ cough**	**cough**	**¬ cough**
covid	**slow**	a1	a2	a3	a4
	medium	a5	a6	a7	a8
	fast	a9	a10	a11	a12
	very fast	a13	a14	a15	a16
¬ covid	**slow**	a17	a18	a19	a20
	medium	a21	a22	a23	a24
	fast	a25	a26	a27	a28
	very fast	a29	a30	a31	a32

- How many entries do I need to store this table? **32 (31 parameters)**

Independent Events

- Total entries: **32 (31 parameters)**

- Now I know that

$$P(Fever, Cough, Covid, I\_Sp) = P(Fever, Cough, Covid)\ P(I\_Sp)$$

- Now I know that
  - To store the above table, I need to store $P(Fever, Cough, Covid)\ and\ P(I\_Sp)$

- Total entries: **32 (31 parameters)**

- Now I know that

$$P(Fever, Cough, Covid, I\_Sp) = P(Fever, Cough, Covid)\ P(I\_Sp)$$

- Now I know that
  - To store the above table, I need to store $P(Fever, Cough, Covid)\ and\ P(I\_Sp)$

Independent Events

	fever		**¬ fever**	
	cough	**¬ cough**	**cough**	**¬ cough**
covid	0.21	0.10	0.11	0.08
¬ covid	0.11	0.07	0.09	0.23

- Table for $P(Fever, Cough, Covid)$
- How many entries: **8 (7 parameters)**

Independent Events

slow	medium	fast	very fast
0.2	0.4	0.25	0.15

- Table for $P(I\_Sp)$
- How many entries? **4 (3 parameters)**

Independent Events

		fever		¬ fever	
		cough	**¬ cough**	**cough**	**¬ cough**
covid	**slow**	a1	a2	a3	a4
	medium	a5	a6	a7	a8
	fast	a9	a10	a11	a12
	very fast	a13	a14	a15	a16
¬ covid	**slow**	a17	a18	a19	a20
	medium	a21	a22	a23	a24
	fast	a25	a26	a27	a28
	very fast	a29	a30	a31	a32

- Total entries required was: **32 (31 parameters)**

- After performing factorization $P(Fever, Cough, Covid, I\_Sp) = P(Fever, Cough, Covid)\ P(I\_Sp)$
  - Total entries required: 8+4= **12 (7+3=10 parameters)**