Lecture 2: Conditioning, Bayes' Rule, and Independence

Yi, Yung (이융)

EE210: Probability and Introductory Random Processes
KAIST EE

August 25, 2021

Roadmap

(1) Conditional Probability
- How should I change my belief about event A, if I come to know that event B occurs?

(2) Bayes' Rule and Bayesian Inference
- prob. of A given that B occurs vs. prob. of B given that A occurs

(3) Independence, Conditional Independence
- Can I ignore my knowledge about event B, when I consider event A?

Roadmap

(1) Conditional Probability

(2) Bayes' Rule and Bayesian Inference

(3) Independence, Conditional Independence

Motivating Example

- Pick a person a at random
  - event A: a's age ≤ 20
  - event B: a is married
- (Q1) What is the probability of A?
- (Q2) What is the probability of A, if I know that that B is true?
- Clearly, the above two should be different. I will assign lower probability for (Q2).
- Question: How should I change my belief, given some additional information?
- Need to build up a new theoretical concept, which we call $\boxed{\text{conditional probability}}$.

Conditional Probability: Notation

- First, let's choose the notation. "Probability of A, given B occurs". What do you recommend?

$$\mathbb{P}(A)(B), \quad \mathbb{P}_B(A), \quad \mathbb{P}^B(A), \quad (B)\mathbb{P}(A), \ldots$$

- People's choice is ... $\boxed{\mathbb{P}(A \mid B)}$
- From now on, given B, $\mathbb{P}(\cdot|B)$ should be a new $\boxed{\text{probability law.}}$
  - Three axioms[1] should be satisfied.

[1] Non-negativity, Normalization, Countable Additivity

Conditional Probability: Definition (1)

- Second, let's define $\mathbb{P}(A|B)$. What would it be a good definition?
- Probability of A given B $\rightarrow$ both A and B occur. Then, what about this?

$$\mathbb{P}(A \mid B) \triangleq \mathbb{P}(A \cap B)$$

- Is it good or bad? Why good? Why bad?
- Reasons why it is bad:
  - $\mathbb{P}(\cdot|B)$ should be a new probability law (thus, three axioms)
  - $\mathbb{P}(\Omega|B) = 1$?
  - $\mathbb{P}(B|B) = 1$ from our common sense.
  - True?

Conditional Probability: Definition (2)

KAIST EE

- How to fix this? Normalization

$$\mathbb{P}(A \mid B) \triangleq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{for} \quad \mathbb{P}(B) > 0.$$

  - Note that this is a definition, not a theorem.
  - So, it's not about right or wrong. It's about how happy we are about this definition.
- All properties of the law $\mathbb{P}(\cdot)$ is applied to the conditional law $\mathbb{P}(\cdot|B)$.
  - Non-negativity. $\mathbb{P}(A|B)$ for any event A?
  - Finite additivity and thus countable additivity. For any two disjoint A and C,

$$\mathbb{P}(A \cup C| \ B) = \frac{\mathbb{P}\Big[(A \cup C) \cap B\Big]}{\mathbb{P}(B)} = \frac{\mathbb{P}\Big[(A \cap B) \cup (C \cap B)\Big]}{\mathbb{P}(B)} = \mathbb{P}(A| \ B) + \mathbb{P}(C| \ B)$$

Roadmap

KAIST EE

(1) Conditional Probability

(2) Bayes' Rule and Bayesian Inference

(3) Independence, Conditional Independence

From now on, using the theory of probability and conditional probability constructed so far, we will develop interesting properties and theorems which are very useful to answer some exciting questions.

That is *Bayes' Rule* to make some *inference (추론).*

Example: (Bayesian) Inference

- A_1: Happy (:-)), A_2: Sad (:-()
- B: Shout
- Assume that somebody gives you the following information:

  $\mathbb{P}(A_1), \quad \mathbb{P}(A_2), \quad \mathbb{P}(B|A_1), \quad \mathbb{P}(B|A_2).$
- Question: $\mathbb{P}(A_1|B)$ and $\mathbb{P}(A_2|B)$?

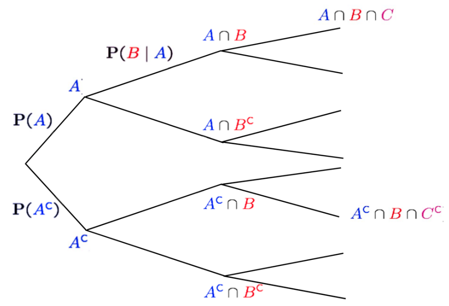
- A_i: state/cause/original value
- B: result/resulting action/noisy measurement
- In reality, $\mathbb{P}(A_i)$ and $\mathbb{P}(B|A_i)$ (cause $\to$ result) can be given from my model
- Inference: $\mathbb{P}(\text{cause} \mid \text{result})$?

We will study this topic rigorously later in this class (chapter 8).

Multiplication Rule

- $\mathbb{P}(B|A) = \boxed{\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}}$
- $\mathbb{P}(A \cap B) = \boxed{\mathbb{P}(A)\mathbb{P}(B|A)}$
- $\mathbb{P}(A^c \cap B \cap C^c) = \boxed{\mathbb{P}(A^c \cap B) \cdot \mathbb{P}(C^c | A^c \cap B)}$

  $= \boxed{\mathbb{P}(A^c) \cdot \mathbb{P}(B|A^c) \cdot \mathbb{P}(C^c | A^c \cap B)}$



Generally,

VIDEO PAUSE

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1, A_2) \cdots \mathbb{P}(A_n|A_1, A_2, \ldots, A_{n-1})$$

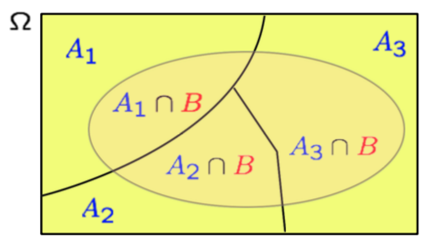
Total Probability Theorem

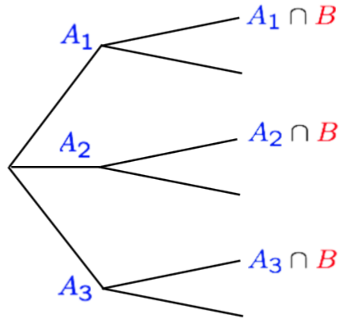
- Partition of Ω into A_1, A_2, A_3
- We know: $\mathbb{P}(A_i)$ and $\mathbb{P}(B|A_i)$
- What is $\mathbb{P}(B)$? (probability of result)

Total Probability Theorem

$$\mathbb{P}(B) = \sum_i \mathbb{P}(A_i)\mathbb{P}(B|A_i)$$

- $\mathbb{P}(A_i \cap B) = \mathbb{P}(A_i)\mathbb{P}(B|A_i)$
- Weighted average from the point of A_i knowledge.





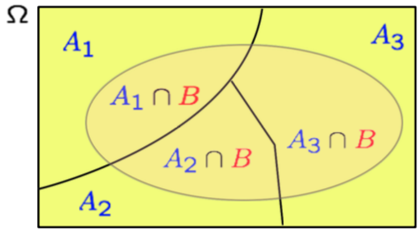
[1] Partition: A_1, A_2, A_3 are mutually exclusive and $\Omega = A_1 \cup A_2 \cup A_3$

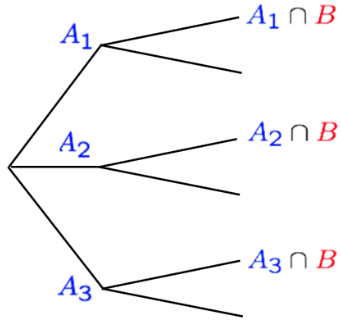
Bayes' Rule

- Partition of Ω into A_1, A_2, A_3
- We know: $\mathbb{P}(A_i)$ and $\mathbb{P}(B|A_i)$
- What is $\mathbb{P}(A_i|B)$?
- revised belief about A_i, given B occurs

Bayes' Rule

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\sum_j \mathbb{P}(A_j)\mathbb{P}(B|A_j)}$$





Example 1: Airplane-Radar

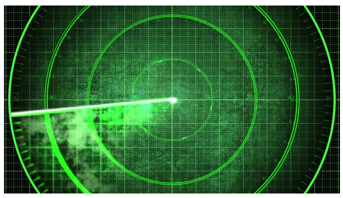
VIDEO PAUSE

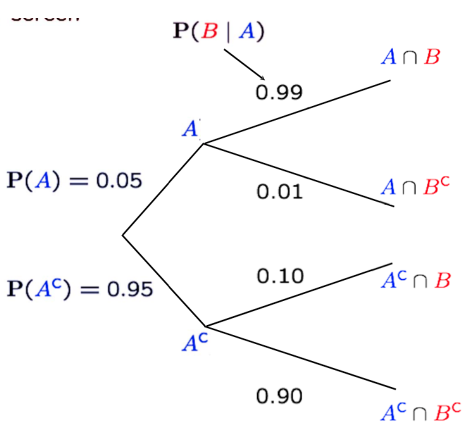
- A : Airplane is flying above
- B : Something on radar screen

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$$
$$= 0.05 \times 0.99 = 0.0495$$

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$$
$$= 0.05 \times 0.99 + 0.95 \times 0.1 = 0.1445$$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0.0495}{0.1445} \approx 0.34$$





Example 2: Happy/Sad-Shout

- A_1: you are happy, A_2: you are sad
- B: you shout.
- Assume:

$$\mathbb{P}(A_1) = 0.7, \ \mathbb{P}(A_2) = 0.3,$$
$$\mathbb{P}(B|A_1) = 0.3, \ \mathbb{P}(B|A_2) = 0.5.$$

- Calculate $\mathbb{P}(A_1|B)$ and $\mathbb{P}(A_2|B)$.

VIDEO PAUSE

$$\mathbb{P}(A_1)\mathbb{P}(B|A_1) = 0.7 \times 0.3 = 0.21$$
$$\mathbb{P}(A_2)\mathbb{P}(B|A_2) = 0.3 \times 0.5 = 0.15$$
$$\mathbb{P}(B) = 0.21 + 0.15 = 0.36$$

$$\mathbb{P}(A_1|B) = \frac{0.21}{0.36} \approx 0.583$$
$$\mathbb{P}(A_2|B) = \frac{0.15}{0.36} \approx 0.417$$

Roadmap

(1) Conditional Probability

(2) Bayes' Rule and Bayesian Inference

(3) Independence, Conditional Independence

Bayesian inference was really fun.

Now, let's develop a new concept from conditioning.

That is *Independence*.

Why We Care Independence?

- Event A: I get the grade A in the probability class (my interest).
- Event B: My friend is rich.

- A and B do not seem dependent on each other. So, just forget B!
- Independence makes our analysis and modeling much simpler, because I can remove independent events in the analysis of what I am interested in.

Independence

- Occurrence of A provides no new information about B. Thus, knowledge about A does NOT change my belief about B.

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

- Using $\mathbb{P}(B|A) = \mathbb{P}(B \cap A)/\mathbb{P}(A)$,

Independence of A and B, $\boxed{A \perp\!\!\!\perp B}$

$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$

- The above definition show the symmetry of independence more clearly.
- Q1. A and B disjoint $\implies A \perp\!\!\!\perp B$?
  No. Actually, really dependent, because if you know that A occurred, then, we know that B did not occur.
- Q2. If $A \perp\!\!\!\perp B$, then $A \perp\!\!\!\perp B^c$? Yes.

Conditional Independence

- Remember: for a probability law $\mathbb{P}(\cdot)$, given some event C, $\mathbb{P}(\cdot|C)$ is a new probability law.
- Thus, we can talk about independence under $\mathbb{P}(\cdot|C)$.
- Given that C occurs, occurrence of A provides no new information about B.

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

- Using $\mathbb{P}(A \cap B|C) = \frac{\mathbb{P}[B \cap (A \cap C)]}{\mathbb{P}(C)} = \frac{\mathbb{P}(A \cap C)\mathbb{P}(B|A \cap C)}{\mathbb{P}(C)} = \mathbb{P}(A|C)\mathbb{P}(B|C)$,

Conditional Independence of A and B given C, $\boxed{A \perp\!\!\!\perp B|C}$

$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \times \mathbb{P}(B|C)$

(Q1) $A \perp\!\!\!\perp B \rightarrow A \perp\!\!\!\perp B|C$?

- Suppose that A and B are independent. If you heard that C occurred, A and B are still independent? VIDEO PAUSE

- Two independent coin tosses
  - H_1: 1st toss is a head
  - H_2: 2nd toss is a head
  - D: two tosses have different results.

- $\mathbb{P}(H_1|D) = 1/2$, $\mathbb{P}(H_2|D) = 1/2$

- $\mathbb{P}(H_1 \cap H_2|D) = 0$,

- No.

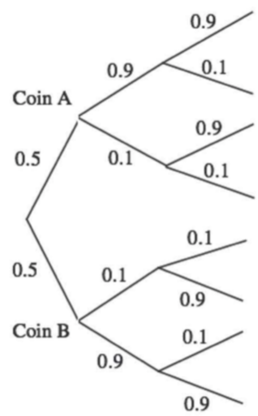
(Q2) $A \perp\!\!\!\perp B|C \rightarrow A \perp\!\!\!\perp B$?

- Two coins: Blue and Red. Choose one uniformly at random, and proceed with two independent tosses.
- $\mathbb{P}(\text{head of blue}) = 0.9$ and $\mathbb{P}(\text{head of red}) = 0.1$
  H_i: i-th toss is head, and B: blue is selected.
- $H_1 \perp\!\!\!\perp H_2|B$? Yes

$$\mathbb{P}(H_1 \cap H_2|B) = 0.9 \times 0.9, \quad \mathbb{P}(H_1|B)\mathbb{P}(H_2|B) = 0.9 \times 0.9$$

- $H_1 \perp\!\!\!\perp H_2$? No

$$\begin{aligned}
\mathbb{P}(H_1) &= \mathbb{P}(B)\mathbb{P}(H_1|B) + \mathbb{P}(B^c)\mathbb{P}(H_1|B^c) \\
&= \frac{1}{2}0.9 + \frac{1}{2}0.1 = \frac{1}{2} \\
\mathbb{P}(H_2) &= \mathbb{P}(H_1) \quad \text{(because of symmetry)} \\
\mathbb{P}(H_1 \cap H_2) &= \mathbb{P}(B)\mathbb{P}(H_1 \cap H_2|B) + \mathbb{P}(B^c)\mathbb{P}(H_1 \cap H_2|B^c) \\
&= \frac{1}{2}(0.9 \times 0.9) + \frac{1}{2}(0.1 \times 0.1) \neq \frac{1}{2}
\end{aligned}$$



Independence of Multiple Events

- Three events: A_1, A_2, A_3. What are the conditions of "their independence"?
- What about this? (Pairwise independence)

  $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2),\ \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_3),\ \mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2)\mathbb{P}(A_3)$
- What about $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$?
- We need both.

Independence of Multiple Events

The events $A_1, A_2, \ldots, A_n$ ar said to be independent if

$$\mathbb{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbb{P}(A_i), \quad \text{for every subset } S \text{ of } \{1, 2, \ldots, n\}$$

Questions?

KAIST EE

1) What is conditional probability? Why do we need it?
2) Explain the overall framework of Bayesian inference.
3) What is the total probability theorem?
4) What is Bayes' rule? What does it can give us?
5) What's the difference between independence and conditional independence?