KAIST EE

Lecture 2: Conditioning, Bayes' Rule, and Independence

Yi, Yung (이융)

EE210: Probability and Introductory Random Processes
KAIST EE

MONTH DAY, 2021

Outline

KAIST EE

- Conditional Probability
- Bayes' Rule
- Bayesian Inference: Sneak Peek
- 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, given that B is true?
- Clearly the above two should be different.
- Question: How should I change my belief, given some additional information?
- Need to build up a new theory, which we call conditional probability.

Conditional Probability: Notation

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

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

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

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?

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

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

Conditional Probability: Definition (2)

- How to fix this? $\boxed{\text{Normalization.}}$

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

  - Note that this is a definition, not a theorem.
- Easily check: Nonnegativity and countable additivity.
- All other properties of the law $\mathbb{P}(\cdot)$ is applied to the conditional law $\mathbb{P}(\cdot|B)$.
- For example, finite additivity. For two disjoint events A and C,

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

Example: Conditional Probability

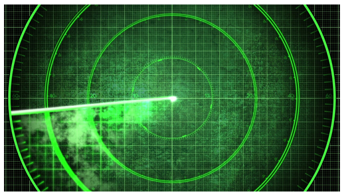
KAIST EE

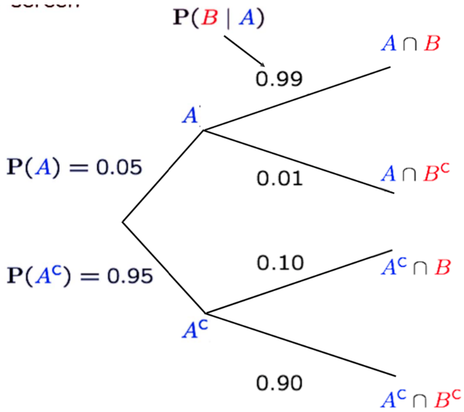
- A : Airplane is flying above
- B : Something on radar screen

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

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

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





KAIST EE

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: you are happy, A_2: you are sad
- B: you 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 $\mathbb{P}(A_2|B)$?

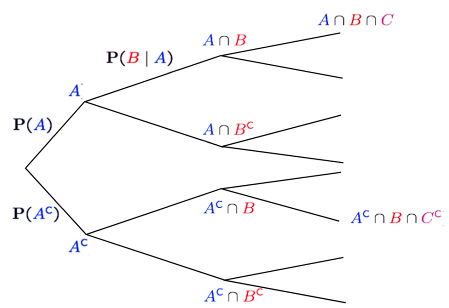
- A_i: state/cause/original value
- B: result/resulting action/noisy measurement
- In reality, $\mathbb{P}(A_i)$ (prior) and $\mathbb{P}(B|A_i)$ (cause $\to$ result) can be given from my model
- Inference: $\mathbb{P}(\text{cause} \mid \text{result})$?

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

Multiplication Rule

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

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



Generally,

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

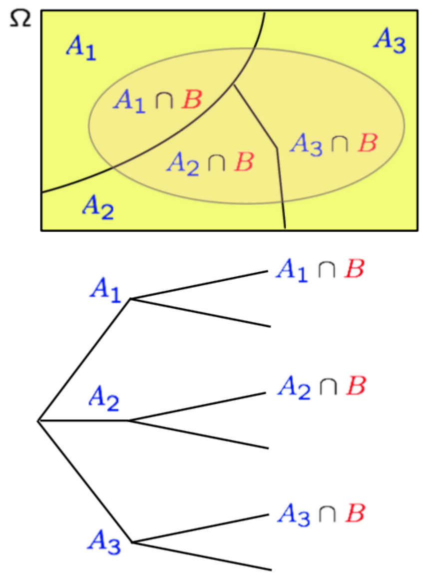
Total Probability Theorem

- Partition of Ω into A_1, A_2, A_3
- We know from my model: $\mathbb{P}(A_i)$ 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.

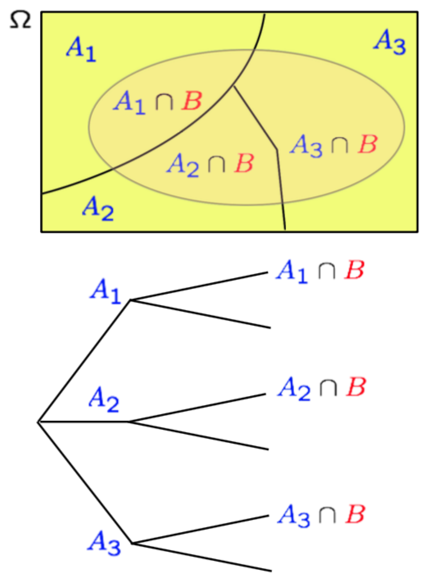


Bayes' Rule

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

Bayes' Rule

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



Bayes' Rule: Example

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

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

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

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 my analysis.

Independence

- Occurrence of A provides no new information about B. Thus, knowledge about A does no 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, $\boxed{A \perp\!\!\!\perp B}$

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

- Q1. A and B disjoint $\Longrightarrow A \perp\!\!\!\perp B$?
  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, say B, $\mathbb{P}(\cdot|B)$ is a new probability law.
- Thus, we can talk about independence under $\mathbb{P}(\cdot|B)$.
- Given that C occurs, occurrence of A provides no new information about B.

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

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. If $A \perp\!\!\!\perp B$, then $A \perp\!\!\!\perp B|C$? Suppose that A and B are independent. If you heard that C occurred, A and B are still independent?
- Q2. If $A \perp\!\!\!\perp B|C$, $A \perp\!\!\!\perp B$?

$A \perp\!\!\!\perp B \rightarrow A \perp\!\!\!\perp B|C$?

- 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.

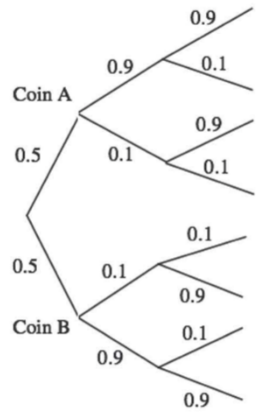
$A \perp\!\!\!\perp B | C \rightarrow A \perp\!\!\!\perp B$?

- 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

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



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

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

Questions?

Review Questions

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?