Lecture 5: Conditional Probability - Part II

Satyajit Thakor
IIT Mandi

Conditional probability

- By the definition of conditional probability,

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A), \quad \text{if } P(A) > 0, P(B) > 0.$$

- For three events A_1, A_2, A_3 with positive probabilities,

$$\begin{aligned} P(\underbrace{A_1, A_2, A_3}) &= P(A_1|A_2, A_3)P(A_2, A_3) \\ &= P(A_1|A_2, A_3)P(A_2|A_3)P(A_3) \end{aligned}$$

Let "∩" as ",".

- Similarly, $P(A_1, A_2, A_3) = P(A_2|A_1, A_3)P(A_1|A_3)P(A_3)$ and in fact, we can have 6 distinct expressions by permuting A_1, A_2, A_3.

– Writing the probability of intersection of events as product of conditional probabilities can be useful to solve problems.

Conditional probability

- Theorem: For events $A_1, \ldots, A_n$ with $P(A_1, A_2, \ldots, A_{n-1}) > 0$,

$$P(A_1, A_2, \ldots, A_n) = P(A_1)P(A_2|A_1)\ldots P(A_n|A_1, \ldots, A_{n-1}).$$

- In fact, we have have $n!$ theorems by permuting $A_1, A_2, \ldots, A_n$.
- Theorem (Bayes' rule):

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

given: $P(A) > 0$, $P(B) > 0$.

- The odds of an event A are $\text{odds}(A) = P(A)/P(A^c)$.
- E.g., if $P(A) = 2/3$ than the odds of A are 2 to 1.
- Note that

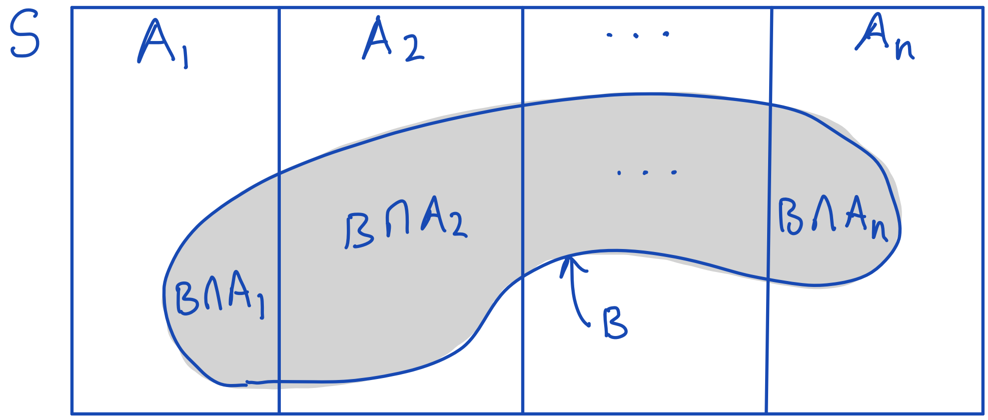
$$P(A) = \text{odds}(A)(1 - P(A))$$
$$\Rightarrow P(A) = \frac{\text{odds}(A)}{1 + \text{odds}(A)}.$$

Conditional probability

- $A_1, \ldots, A_n$ is a partition of set S if $A_i \cap A_j = \emptyset, \forall i \neq j$ (i.e., disjoint sets) and $A_1 \cup \ldots \cup A_n = S$.
- Theorem (Law of total probability LOTP): Let $A_1, \ldots, A_n$ be a partition of the sample space S with $P(A_i) > 0$ for all i. Then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

for any event B.



Conditional probability

Proof: $P(B) = P(B \cap S)$

$= P(B \cap (A_1 \cup A_2 \cup \cdots \cup A_n))$

$= P(\underbrace{(B \cap A_1)} \cup \underbrace{(B \cap A_2)} \cup \cdots \cup \underbrace{(B \cap A_n)})$ ($\because$ distributive property of "$\cap$")

disjoint events since A_i's are disjoint

$= P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n)$

($\because$ Axiom 2)

$= P(B|A_1) \cdot P(A_1) + \cdots + P(B|A_n) \cdot P(A_n)$

$$= \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

Conditional probability

- Example: You have one fair coin, and one biased coin which lands Heads with probability $3/4$. You pick one of the coins at random and flip it three times. It lands Heads all three times. Given this information, what is the probability that the coin you picked is the fair one?

A: the event that the chosen coin lands 3 H's.

F: the event that the chosen coin is fair.

we want to find $P(F|A)$.

- It is not easy to find $P(F|A)$ directly.
- It is easier to find $P(A|F)$, $P(A|F^c)$.

Conditional probability

The only out come favorable to the event A is (H,H,H). (out of 8 total outcomes)

By Bayes' rule,

$$P(F|A) = \frac{P(A|F)\cdot P(F)}{P(A)} = \frac{\overbrace{P(A|F)}\cdot P(F)}{\underbrace{P(A|F)\cdot P(F) + P(A|F^c)P(F^c)}_{(\because \text{LOTP})}}$$

$$= \frac{(1/2)^3\cdot(1/2)}{(1/2)^3\cdot(1/2)+(3/4)^3\cdot(1/2)}$$

$$\approx 0.23.$$

- I.e., probability that the coin is fair given that the first 3 are heads is less than the probability that the coin is biased given that the first 3 are heads:

$$P(F^c|A) = 1 - P(F|A)$$

$$\approx 0.77 \quad (\because P(S|A) = P(F|A) + P(F^c|A))$$

Conditional probability

- **Theorem:** Conditional probabilities given an "evidence" event E are all probabilities.

In other words, $P(\cdot \mid E)$ is a valid probability f^n.

Proof: – We need to check whether the axioms are satisfied by $P(\cdot \mid E)$.

– First, note that $P(\cdot \mid E)$ is a f^n from subsets of S to $[0,1]$.

Axiom 1:

$$P(\phi \mid E) = \frac{P(\phi \cap E)}{P(E)} = \frac{P(\phi)}{P(E)} = 0.$$

$$P(S \mid E) = \frac{P(S \cap E)}{P(E)} = \frac{P(E)}{P(E)} = 1.$$

That is, $P(\cdot \mid E)$ satisfies Axiom 1.

Conditional probability

Axiom 2 : Let $A_1, A_2, \ldots, A_n$ be disjoint events. Then,

$$P((A_1 \cup A_2 \cup \cdots \cup A_n) \mid E)$$

$$= \frac{P((A_1 \cup A_2 \cup \cdots \cup A_n) \cap E)}{P(E)}$$

$$= \frac{P((A_1 \cap E) \cup (A_2 \cap E) \cup \cdots \cup (A_n \cap E))}{P(E)}$$

$$= \sum_{i=1}^{n} \frac{P(A_i \cap E)}{P(E)}$$

$$= \sum_{i=1}^{n} P(A_i \mid E).$$

That is, $P(\cdot \mid E)$ also satisfies Axiom 2.

Conditional probability

- Theorem (Bayes' rule with extra conditioning): If $P(A, E) > 0$ and $P(B, E) > 0$ then

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

<u>Proof</u>:

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

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

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

Conditional probability

- Theorem (LOTP with extra conditioning): Let $A_1, \ldots, A_n$ be a partition of S with $P(A_i, E) > 0$ for all i. Then

$$P(B|E) = \sum_{i=1}^{n} P(B|A_i, E)P(A_i|E).$$

Proof:

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

$$= \frac{P(B \cap E \cap (A_1 \cup A_2 \cup \cdots \cup A_n))}{P(E)}$$

$$= \sum_{i=1}^{n} \frac{P(B \cap E \cap A_i)}{P(E)}$$

$$= \sum_{i=1}^{n} P(B|A_i, E) \cdot \underbrace{\frac{P(A_i, E)}{P(E)}}_{= P(A_i|E).}$$

Conditional probability

- Example: Continuing with the "coin" example, suppose that we have now seen our chosen coin land Heads three times. If we toss the coin a fourth time, what is the probability that it will land Heads once more?

Recall: A: the event that the chosen coin lands 3 H's.
F: the event that the chosen coin is fair.

Let: B: the event that the chosen coin lands 4th H's.

- We want to find $P(B|A)$.

$$P(B|A) = P(B|A,F) \cdot P(F|A) + P(B|A,F^c) \cdot P(F^c|A)$$

$$\approx \frac{1}{2}(0.23) + \frac{3}{4}(1-0.23)$$

$$\approx 0.69.$$

Conditional probability

Intuitive Example:

assume that it occured during 1-3 PM.

say A

- Let G be the event that a certain individual is guilty of a certain robbery. In gathering evidence, it is learned that an event E_1 occurred, and a little later it is also learned that another event $E2$ also occurred.
  Is it possible that individually, these pieces of evidence increase the chance of guilt (so $P(G|E_1) > P(G)$ and $P(G|E_2) > P(G)$), but together they decrease the chance of guilt (so $P(G|E_1, E_2) < P(G)$)?

Let E_1: A was in a nearby restaurant during 1-2 PM.

Let E_2: A was in a nearby restaurant during 2-3 PM.

- $P(G)$ may be taken as 1/ Population of the town.

- Note: $P(G|E_i) > P(G)$, $i = 1, 2$. (near the incident $\Rightarrow$ more prob.)

- But: $P(G|E_1, E_2) = 0 < P(G)$.