STA 130A: Mathematical Statistics: Brief Course

Lecture 3: Conditional Probability and Bayes' Rule

Dogyoon Song

Fall 2026, UC Davis

Agenda

Last time:

- Set theory basics
- Probabilistic models

Today:

- Complete Lecture 2
  - Discrete versus continuous models
  - Modeling versus analysis
- Conditional probability
- Law of total probability
- Bayes' rule
- (Time permitting) A first look at independence

Recap: Outcomes, events, probability

Sample space: Ω is the set of possible outcomes; $\omega \in \Omega$ is one outcome

Event: $A \subseteq \Omega$ occurs when $\omega \in A$

- For a die roll, $A = \{2, 4, 6\}$ means "the result is even"
- $A \cap B$: both occur; $A \cup B$: at least one occurs; A^c: A does not occur

Probability law: a function P that assigns each event A a number $P(A)$ such that

- $P(A) \geq 0$
- If $A_1, A_2, \ldots$ are pairwise disjoint, then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$
- $P(\Omega) = 1$

Useful probability rules (derived from the axioms):

- $P(\Omega) = 1, \qquad P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- For finite equally likely outcomes, $P(A) = |A|/|\Omega|$

Completing Lecture 2: Discrete vs. continuous models

In a discrete model, the probabilities of singleton events determine the probability law:

$$P(A) = \sum_{a \in A} P(\{a\}).$$

In a continuous model, singleton probabilities typically do *not* determine the law

Example (A wheel of fortune)

Experiment: Spin a fair wheel and record a number in $[0, 1]$

Sample space: $\Omega = [0, 1]$

Events: subsets of $[0, 1]$

Q: What is the probability $P(\{x\})$ for a single point $x \in [0, 1]$?

- Under the natural uniform model on $[0, 1]$, $P(\{x\}) = 0$ for every x
- Yet intervals can still have positive probability; $P([a, b]) = b - a$ for $0 \leq a \leq b \leq 1$

* Countable additivity does not apply to an uncountable union of singletons

Completing Lecture 2: Models vs. reality

Using the framework of probability theory involves two distinct stages:

1. **Modeling:** Choose a sample space and a probability law to represent reality
   - The axioms ensure mathematical consistency; suitability depends on the situation
   - Different reasonable models may represent the same situation differently
   - Even a simplified "incorrect" model may be useful due to simplicity and tractability
2. **Analysis:** Once the model is fixed, compute probabilities and draw conclusions (*within the model*)
   - This stage is governed by the axioms of probability and ordinary logic
   - The main difficulty is often computation, not ambiguity

Different modeling assumptions can give different probability calculations

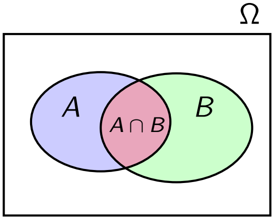
- They do not imply inconsistencies in probability theory $\rightarrow$ see Bertrand's paradox in Lecture 2

Conditional probability: Definition

For an event[1] B with $P(B) > 0$, the **conditional probability** of A given B is

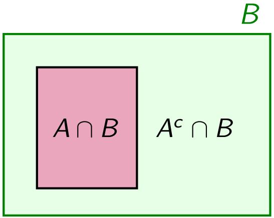
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- **Example (fair die):** $A = \{6\}$, $B = \{2, 4, 6\}$ $\quad \rightarrow \quad$ $P(A) = 1/6$ vs. $P(A \mid B) = 1/3$
- Conditioning on B means we restrict attention to outcomes in B, and renormalize



Before conditioning

$\Longrightarrow$
condition on B



After conditioning on B

[1] In this course, we only condition on events with positive probability

Conditional probability is a probability law

For any fixed B with $P(B) > 0$, the map $P(\cdot \mid B) : A \mapsto P(A \mid B)$ satisfies the axioms

- **Nonnegativity:** $P(A \mid B) \geq 0$ for all events A $\qquad \because P(A \cap B) \geq 0, \quad \forall A$
- **Additivity**[2]**:** If $A_1 \cap A_2 = \emptyset$, then

$$\begin{aligned}
P(A_1 \cup A_2 \mid B) &= \frac{P\big((A_1 \cup A_2) \cap B\big)}{P(B)} && \text{by definition} \\
&= \frac{P\big((A_1 \cap B) \cup (A_2 \cap B)\big)}{P(B)} && \text{by distributivity} \\
&= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} && \text{by additivity of } P \\
&= P(A_1 \mid B) + P(A_2 \mid B) && \text{by definition}
\end{aligned}$$

- **Normalization:** $P(\Omega \mid B) = 1$ $\qquad \because P(\Omega \cap B) = P(B)$

$\implies$ The map $A \mapsto P(A \mid B)$ is a probability law

[2] Countable additivity follows by the same argument

Conditional probability: A coin-toss example

Example (3 fair coin tosses)

Suppose we toss a fair coin three times and let

$$A = \{\text{more heads than tails}\}, \qquad B = \{\text{1st toss is a head}\}$$

Observe that

$$\begin{aligned} \Omega &= \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}, \\ A &= \{HHH, HHT, HTH, THH\}, \\ B &= \{HHH, HHT, HTH, HTT\} \end{aligned}$$

The event $A \cap B = \{HHH, HHT, HTH\}$

Since all 8 outcomes are equally likely,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{3/8}{4/8} = \frac{3}{4} \qquad \text{whereas} \qquad P(A) = \frac{4}{8} = \frac{1}{2}$$

Modeling with conditional probability

Example (Fire alarm)

Define two events

$$F = \{\text{there is a fire}\}, \qquad A = \{\text{the alarm goes off}\}$$

The alarm may go off when there is a fire, but it may also produce a false alarm

Thus, there are four possible scenarios:

F	A	Fire and alarm
	A^c	Missed fire
F^c	A	False alarm
	A^c	Nothing happens

For $P(F) > 0$, **multiply along a case:** $P(F \cap A) = P(F)P(A \mid F)$

Conditional probability: Multiplicative rule

Multiplicative rule

Assume $P(A_1 \cap \cdots \cap A_k) > 0$ for $k = 1, \ldots, n-1$:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \cdots P\left(A_n \,\middle|\, \bigcap_{i=1}^{n-1} A_i\right)$$

This follows from the definition of conditional probability:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot \underbrace{\frac{P(A_2 \cap A_1)}{P(A_1)}}_{=P(A_1 \cap A_2 \mid A_1)} \cdot \underbrace{\frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}}_{=P(A_1 \cap A_2 \cap A_3 \mid A_1 \cap A_2)} \cdots \underbrace{\frac{P\left(\bigcap_{i=1}^{n} A_i\right)}{P\left(\bigcap_{i=1}^{n-1} A_i\right)}}_{=P\left(\bigcap_{i=1}^{n} A_i \,\middle|\, \bigcap_{i=1}^{n-1} A_i\right)}$$

In words, the probability of a sequence of events is the product of conditional probabilities

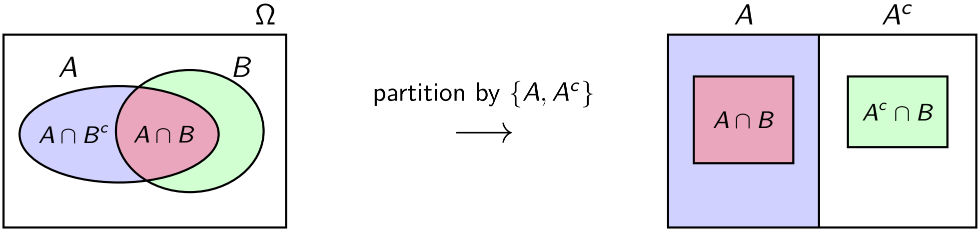
Conditional probability: Law of total probability

Law of total probability

Let $\{A_1, \ldots, A_n\}$ be a partition of Ω such that $P(A_i) > 0$ for all i
Then for any event B,

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

For the partition $\{A, A^c\}$ with $0 < P(A) < 1$, $P(B) = P(A)P(B \mid A) + P(A^c)P(B \mid A^c)$



Law of total probability: Example 1

Example (One match against a random opponent)

You enter a tournament where there are three types of players and let

$$\begin{aligned} A_i &= \text{event of playing with an opponent of type } i, \qquad i \in \{1,2,3\}, \\ B &= \text{event of winning} \end{aligned}$$

Your chance of meeting each player type:

$$P(A_1) = 0.6, \qquad P(A_2) = 0.3, \qquad P(A_3) = 0.1$$

Your chance of winning against each type:

$$P(B \mid A_1) = 0.9, \qquad P(B \mid A_2) = 0.7, \qquad P(B \mid A_3) = 0.5$$

Question: What is your chance of winning a match (against a randomly chosen opponent)?

Answer: $P(B) = \sum_{i=1}^{3} P(A_i)P(B \mid A_i) = 0.6 \times 0.9 + 0.3 \times 0.7 + 0.1 \times 0.5 = 0.80$

Law of total probability: Example 2

Example (The Monty Hall problem)

A prize is equally likely to be found behind any of three closed doors

1. You point to one of the doors
2. A friend, who knows where the prize is, opens one of the other two doors and always reveals no prize
3. You may either stay with your original choice or switch to the other unopened door
4. You win the prize if it lies behind your final choice

Question: Should you stay or switch?

Answer: You should switch

Let $C = \{\text{your initial choice is correct}\}$. Then $P(C) = 1/3$, so $P(C^c) = 2/3$

- If C occurs, then switching loses
- If C^c occurs, the friend is forced to open the only remaining door with no prize, so switching wins

$$P(\text{win by staying}) = P(C) = \frac{1}{3}, \qquad P(\text{win by switching}) = P(C^c) = \frac{2}{3}$$

Bayes' rule

Often, we know $P(B \mid A)$, but want $P(A \mid B)$

- A: a hypothesis/model/state of the world
- B: observed data or evidence
- Example: $A =$ having a disease, $\quad B =$ positive screening result

After observing B, we update $P(A)$ to $P(A \mid B)$

- *Prior:* $P(A)$ and $P(A^c)$
- *Likelihood:* $P(B \mid A)$ and $P(B \mid A^c)$
- *Posterior:* $P(A \mid B)$ and $P(A^c \mid B)$

For B with $P(B) > 0$, **Bayes' rule** states that

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \qquad \text{where} \qquad P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c)$$

- *Posterior* $=$ *Likelihood* $\times$ *Prior* / *Evidence*

Bayes' rule: Example 1

Example (Disease screening)

Let A = disease and B = positive screening result

- Suppose $P(A) = 0.01$
- $P(B \mid A) = 0.8$ (true positive)
- $P(B \mid A^c) = 0.1$ (false positive; the test is positive despite the non-existence of disease)

Q: If the test is positive, how likely is it that the person actually has the disease?

$$\begin{aligned} P(A \mid B) &= \frac{P(B \mid A)\, P(A)}{P(B \mid A)\, P(A) + P(B \mid A^c)\, P(A^c)} \\ &= \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.1 \times 0.99} \approx 0.075 \end{aligned}$$

- A positive result raises the probability from 1% to about 7.5%, but it is still far from certain
- With a low prior probability, false positives can outnumber true positives
  (When the disease is rare, even a fairly accurate test can produce many false positives)

Bayes' rule: Example 2

Example (Fair or double-headed?)

Let A = the coin is double-headed and A^c = the coin is fair

Let B = the first toss is Head

- Suppose $P(A) = 0.5$
- $P(B \mid A) = 1$ (the double-headed (biased) coin always lands Head)
- $P(B \mid A^c) = 0.5$ (the fair coin lands Head with probability $1/2$)

Q: After observing one Head, how likely is it that the coin is double-headed?

$$
\begin{aligned}
P(A \mid B) &= \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)} \\
&= \frac{1 \times 0.5}{1 \times 0.5 + 0.5 \times 0.5} = \frac{2}{3} \approx 0.667.
\end{aligned}
$$

- One Head raises the probability from 50% to 66.7%
- Data increase the probability of the model that better explains the observation

A first look at independence

Sometimes knowing that B occurred does not change the probability of A

- When $P(B) > 0$, this means $P(A \mid B) = P(A)$

Definition (Independence)

Events A and B are **independent** if

$$P(A \cap B) = P(A)P(B)$$

Example: For two independent fair coin tosses, let $A = \{\text{first toss is H}\}$ and $B = \{\text{second toss is H}\}$. Then $P(A \cap B) = 1/4 = (1/2)(1/2)$.

Not the same as disjointness: If $A \cap B = \emptyset$ and both probabilities are positive,

$$0 = P(A \cap B) < P(A)P(B) \qquad \Longrightarrow \qquad \text{A and B are not independent}$$

Wrap-up

Conditional probability: Restrict attention and renormalize

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0$$

- Multiplicative rule: Multiply conditional probabilities along a sequence of outcomes
- Law of total probability: Add over a partition to obtain the total probability

Bayes' rule: Reverse the direction of conditioning

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

* Posterior = likelihood × prior / evidence

Independence: $P(A \cap B) = P(A)P(B)$

Suggested reading: [BT08, Ch. 1.3 & 1.4]

Next lecture: Independence, independent trials, and counting

References

Dimitri Bertsekas and John N Tsitsiklis.
Introduction to probability, volume 1.
Athena Scientific, 2nd edition, 2008.