

# **Section 2: Probability**

STA 35C – Statistical Data Science III

**Instructor:** Akira Horiguchi

Fall Quarter 2025 (Sep 24 – Dec 12)  
MWF, 12:10 PM – 1:00 PM, Olson 158  
University of California, Davis

## Section 1: Overview

Based on Chapter 1 of textbook: <https://www.probabilitycourse.com/>

### 1 Probability

- Finding probabilities
- Probability models: discrete vs continuous

### 2 Conditional probability

- Independence
- Law of total probability
- Bayes' rule
- Conditional independence

# Probability

## Probability measure: introduction

Probability is a way to quantify randomness and/or uncertainty.

- e.g., coin flips, dice rolls, stocks, weather.
- Rules of probability should be intuitive and self-consistent.
- Self-consistent: the rules shouldn't lead to contradictions.
- Thus these rules must be constructed in a certain way.
- Suppose we want to assign a probability to each event in a set of possible events.
- We would like, at the very least:
  1. each probability to be a value between 0 and 1 (inclusive)
  2. the probability assigned to the full set of events to be 1
    - close to 1  $\Rightarrow$  very likely that A occurs.
  3. the probability assigned to the empty set to be 0
    - close to 0  $\Rightarrow$  very unlikely that A occurs.
- We need more restrictions to ensure self-consistency.

The following definition will lead to intuitive and self-consistent rules of probability.

- We assign a *probability* measure  $P(A)$  to an event A.

## Probability measure: definition

### Definition 1: Probability measure $P(\cdot)$

For a nonempty sample space  $\Omega$ , the set function  $P: \Omega \rightarrow [0, 1]$  is a *probability measure*, if

- $P(\Omega) = 1$ ,
- for any pairwise disjoint events  $A_1, A_2, A_3, \dots \subset \Omega$  (i.e.  $A_i \cap A_j = \emptyset$  for all  $i, j$  with  $i \neq j$ ), holds:

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots \quad (1)$$

This definition fulfills the three desirable properties:

- $P(\Omega) = 1$ : the probability of the biggest possible set is equal to 1.
- Property (1) – called the *countable additivity* property – allows us to add probabilities of disjoint sets.

# Probability

## Finding probabilities

Given a random experiment with a sample space  $\Omega$ , how do we find the probability of an event of interest? Use:

- the specific information that we have about the random experiment.
- the probability rules induced by Definition 1.

## Finding probabilities: example

Example: Roll a fair four-sided die. What is the probability of  $E = \{1, 3\}$ ?

- Information about experiment (fair die):  $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\})$ .
- Probability rules:

$$\begin{aligned}1 &= P(S) \\&= P(\{1\} \cup \{2\} \cup \{3\} \cup \{4\}) \\&= P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) \\&= 4P(\{1\}).\end{aligned}$$

Thus  $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = \frac{1}{4}$ . Finally,

$$P(E) = P(\{1, 3\}) = P(\{1\}) + P(\{3\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

## Finding probabilities: notation

Annoying to write e.g.,  $P(\{2\})$

- Simplify to  $P(2)$
- But always keep in mind that  $P$  is a function on sets, not on individual outcomes.

## Finding probabilities: more tools

Definition 1 implies the following additional properties:

### Properties of $P(\cdot)$

Given a sample space  $\Omega$  and arbitrary events  $A, B \subset \Omega$ , Definition 1 implies

1.  $P(\emptyset) = 0$
2.  $P(A^c) = 1 - P(A)$
3.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4.  $P(B \setminus A) = P(B) - P(A \cap B)$
5.  $P(A) \leq P(B)$  if  $A \subset B$ .

(Pictures for intuition; for formal proofs, see “Example 1.10” in §1.3.3 of textbook)

## Finding probabilities: example

Suppose we have the following information:

1. There is a 60 percent chance that it will rain today.
2. There is a 50 percent chance that it will rain tomorrow.
3. There is a 30 percent chance that it does not rain either day.

Find the following probabilities:

- a. The probability that it will rain today or tomorrow.
- b. The probability that it will rain today and tomorrow.
- c. The probability that it will rain today but not tomorrow.
- d. The probability that it either will rain today or tomorrow, but not both.

# Probability

Probability models: discrete vs continuous

## Probability models: discrete vs continuous

Distinguish between two different types of sample spaces: *discrete* and *continuous*.

- Will discuss in more detail in Section 3 of the course.
- Discrete: can compute the probability of an event by adding all outcomes in the event.
- Continuous: need to use integration instead of summation.

## Probability models: discrete

If a sample space  $\Omega$  is a countable set, this refers to a *discrete* probability model.

- Can list all elements:  $\Omega = \{s_1, s_2, s_3, \dots\}$ .
- For an event  $A \subset \Omega$ , by countable additivity (1) we can write

$$P(A) = P\left(\bigcup_{s \in A} \{s\}\right) = \sum_{s \in A} P(s)$$

Thus, to find probability of an event, just need to sum the probability of individual elements in that event.

## Probability models: discrete (example)

Consider a gambling game: win  $k - 2$  dollars with probability  $\frac{1}{2^k}$  for any  $k \in \mathbb{N}$ .

- What is the probability of winning at least \$1 and less than \$4?
- What is the probability of winning more than \$1?

## Probability models: discrete (equally likely outcomes)

Important special case: finite sample space  $\Omega$  where each outcome is equally likely.

- Thus for any outcome  $s \in \Omega$ , we must have

$$P(s) = \frac{1}{|\Omega|}.$$

- In such a case, for any event  $A$ , we can write

$$P(A) = \sum_{s \in A} P(s) = \sum_{s \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}.$$

## Probability models: continuous

Consider a sample space  $\Omega$  that is an *uncountable* set.

- E.g., a 50-minute exam (so  $\Omega = [0, 50]$ ), and let  $T_{Ant}$  be the time it takes Ant to finish the exam.
- What is the probability of  $T_{Ant} \in [42.5, 45)$ ?

# Conditional probability

## Introduction

As you obtain *additional information*, how should you update probabilities of events?

- For example, suppose I roll a fair die.
- Let  $A = \{1, 3, 5\}$ . What is the probability that the outcome is in  $A$ ?  
We will write this as  $P(A)$ .
  
- Let  $B = \{1, 2, 3\}$ . What is the probability of  $A$  if I know that the outcome is in  $B$ ?  
We will write this as  $P(A|B)$ .

In the previous example, we call...

- ... $P(A)$  the *prior probability* of  $A$ ;
- ... $P(A|B)$  the *conditional probability of  $A$  given that  $B$  has occurred*.
  - ▶ Usually shortened to the *conditional probability of  $A$  given  $B$* .

The way we obtained  $P(A|B)$  in this example can be generalized by the following definition.

## Definition

### Definition 2: Conditional probability

If  $A$  and  $B$  are two events in a sample space  $\Omega$ , then the *conditional probability of  $A$  given  $B$*  is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ when } P(B) > 0. \quad (2)$$

If we know that  $B$  has occurred, then we can discard every outcome outside of  $B$ .

- In other words, our sample space is reduced to the set  $B$ . (Picture)
  
- We divide  $P(A \cap B)$  by  $P(B)$  so that the conditional probability of the new sample space  $B$  becomes 1.
- $P(A|B)$  is undefined when  $P(B) = 0$  (meaning  $B$  never occurs).

## Conditional probability rules

Conditional probability itself is a probability measure.

- So all probability rules learned so far can be extended to conditional probability.

For example, Definition 1 (slide 3) and other properties (slide 7)

## Important special cases

Plug into Definition 2

- When A and B are disjoint:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0$$

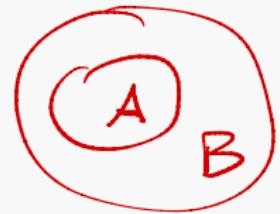
- When B is a subset of A:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$



- When A is a subset of B:

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



## Example

I roll a fair die twice. Let

- $X_1$  be the result of the first roll;
- $X_2$  be the result of the second roll;

Given that I know  $X_1 + X_2 = 7$ , what is the probability that  $X_1 = 4$  or  $X_2 = 4$ ?

- Let  $B$  be the event that  $X_1 + X_2 = 7$ . Let  $A$  be the event that  $X_1 = 4$  or  $X_2 = 4$ .

$$B = \{(1, 6), (2, 5), (3, 4), \underline{(4, 3)}, \underline{(5, 2)}, (6, 1)\}$$

$$A \cap B = \{(3, 4), (4, 3)\}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \stackrel{\downarrow}{=} \quad \frac{|A \cap B| / 36}{|B| / 36} = \frac{2}{6}$$

## Chain rule

We can rearrange the formula in Definition 2 as

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

We can generalize this to 3 events:

$$P(A \cap B \cap C) = \underbrace{P(A)}_{P(A)} \underbrace{P(B \cap C | A)}_{P(B|A) \quad P(C|B, A)}$$

$$P(\underbrace{B \cap C}_{\substack{| \\ A}}) = P(\underbrace{\quad}_{|A}) P(\underbrace{\quad}_{|A})$$

We can generalize this to  $n \geq 2$  events (chain rule for conditional probability):

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_2, A_1) \cdots P(A_n | A_{n-1}, A_{n-2} \cdots A_1)$$

## Example

Of 100 units of a certain product, 5 are defective. If we pick three of the 100 units at random, what is the probability that none of them are defective?

- For  $i = 1, 2, 3$ , let  $A_i$  be the event that the  $i$ th picked unit is NOT defective.

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \underbrace{P(A_2 | A_1)}_{\frac{94}{99}} \underbrace{P(A_3 | A_2, A_1)}_{\frac{93}{98}}$$

$$P(A) \quad P(B|A) \quad P(C|B, A)$$

# Conditional probability

## Independence

## Introduction

Let  $A$  be the event that it rains tomorrow. Let  $B$  be the event that the coin I toss (indoors) tomorrow lands heads up.

- Should the result of the coin toss depend on tomorrow's weather?
- Should the probability of  $A$  depend on whether or not  $B$  happens?
- Two events are *independent* if one does not convey any info about the other.

## Definition

### Definition 3: Independent events

Two events  $A$  and  $B$  are *independent* if and only if

$$P(A \cap B) = P(A)P(B). \quad (3)$$

If two events  $A$  and  $B$  are independent and  $P(B) \neq 0$ , then

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

i.e., the conditional probability  $P(A|B)$  is the same as the prior probability  $P(A)$ .

- Sometimes it is obvious if two given events are independent or not.
- Other times, we need to check if they satisfy the independence condition ~~(3)~~.

l3>

## Example

I pick a random number from  $\{1, 2, 3, \dots, 10\}$ , and call it  $N$ .

- Suppose that all outcomes are equally likely.
- Let  $A$  be the event that  $N < 7$ , and let  $B$  be the event that  $N$  is even.

Are  $A$  and  $B$  independent?

$$P(A) = \frac{6}{10} \quad P(B) = \frac{5}{10}$$
$$A \cap B = \{2, 4, 6\} \quad \Rightarrow \quad P(A \cap B) = \frac{3}{10}$$

$\therefore A$  and  $B$  are independent

## Corollary

The following result can now be proven:

### Corollary 1

If events  $A$  and  $B$  are independent, then

- $A$  and  $B^c$  are independent,
- $A^c$  and  $B$  are independent, and
- $A^c$  and  $B^c$  are independent.

## Independent $\neq$ disjoint

- Two *independent* events convey no information about the other.
- Two *disjoint* events cannot occur at the same time.

## Definition: extend to $\geq 2$ events

### Definition 4: Independent events ( $\geq 2$ )

For  $n \geq 2$ , events  $A_1, A_2, \dots, A_n$  are *independent* if and only if we have

$$P\left(\bigcap_{k \in \mathcal{K}} A_k\right) = \prod_{k \in \mathcal{K}} P(A_k). \quad \leftarrow \quad (4)$$

for every nonempty subset  $\mathcal{K} \subset \{1, 2, \dots, n\}$ .

If  $n=3$ , need to check

$$P(A_1 \cap A_2) = P(A_1) P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1) P(A_3)$$

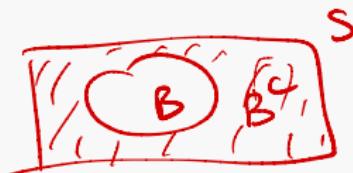
$$P(A_2 \cap A_3) = P(A_2) P(A_3)$$

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

## **Conditional probability**

**Law of total probability**

# Result



## Law of Total Probability:

If events  $B_1, B_2, B_3, \dots$  form a partition of the sample space, then for any event  $A$  we have

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i) P(B_i). \quad (5)$$

Because  $B$  and  $B^c$  partition the sample space,  
from (5) we get:

$$P(A) = P(A|B) P(B) + P(A|B^c) P(B^c).$$

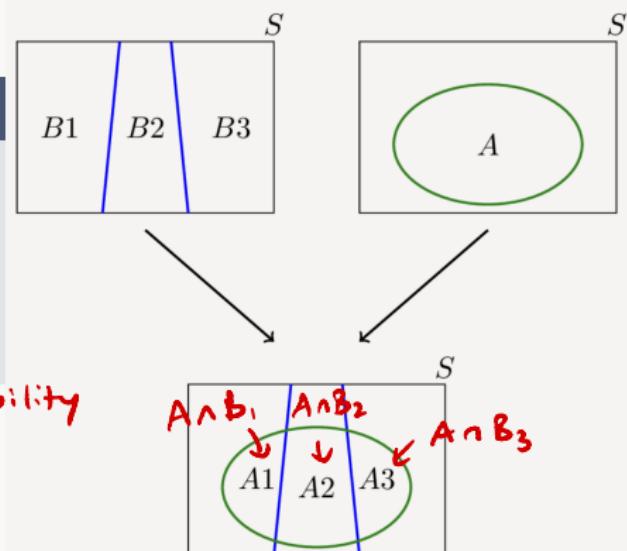


Fig.1.24 - Law of total probability.

$$A = \underbrace{(A \cap B_1)}_{\text{disjoint}} \cup \underbrace{(A \cap B_2)}_{\text{disjoint}} \cup \underbrace{(A \cap B_3)}_{\text{disjoint}}$$

## An example

Suppose there is a population of students who are left- or right-handed (assume that no student is ambidextrous). We know that:

- 30% of these students are taller than 6 feet, and of these, 40% are left-handed.
- Of the remaining 70% of students, 20% are left-handed.

Using the law of total probability, calculate the probability that a student chosen uniformly at random from this population is left-handed.

$$\begin{aligned} \bullet P(> 6 \text{ ft}) &= \underline{\underline{0.3}} \\ \bullet P(L | > 6 \text{ ft}) &= \underline{\underline{0.4}} \\ \bullet P(L | \leq 6 \text{ ft}) &= \underline{\underline{0.2}} \end{aligned}$$

$$P(L) = \underbrace{P(L | > 6 \text{ ft})}_{0.4} \underbrace{P(> 6 \text{ ft})}_{0.3} + \underbrace{P(L | \leq 6 \text{ ft})}_{0.2} \underbrace{P(\leq 6 \text{ ft})}_{1 - 0.3}$$

# Conditional probability

Bayes' rule

# Introduction

From the definition of conditional probability, we know for any two events  $A$  and  $B$  that

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

Dividing by  $P(A)$  (assuming it is not zero), we get *Bayes' rule*:

$$\underline{P(B|A)} = \frac{P(A|B) P(B)}{P(A)} \quad (6)$$

Often  $P(A)$  is unknown and difficult to deduce.

- Sometimes we can use the law of total probability (5).

## Example: False positive paradox

$$P(D) = \frac{1}{10000} = 0.0001$$

A certain disease affects about 1 out of 10,000 people. There is a test to check whether the person has the disease. In particular, we know that

- the probability that the test result is positive, given that the person does not have the disease, is 2%;  $\rightarrow P(+|N) = 0.02$

- the probability that the test result is negative, given that the person has the disease, is 1%.  $\rightarrow P(-|D) = 0.01 \Rightarrow P(+|D) = 1 - 0.01 = 0.99$

Suppose a random person gets tested for the disease and the test result is positive.

What is the probability that the person has the disease?

$$P(D|+) = \frac{P(+|D) P(D)}{P(+)} \approx \frac{0.99 \cdot 0.0001}{0.02} \approx 0.005$$

$$\begin{aligned} P(+) &= P(+|D) P(D) + P(+|N) P(N) \\ &= \frac{0.99}{0.0001} + \frac{0.02}{0.9999} \\ &= 0.000099 + 0.019998 \\ &\approx 0.02 \end{aligned}$$

## Example: False positive paradox

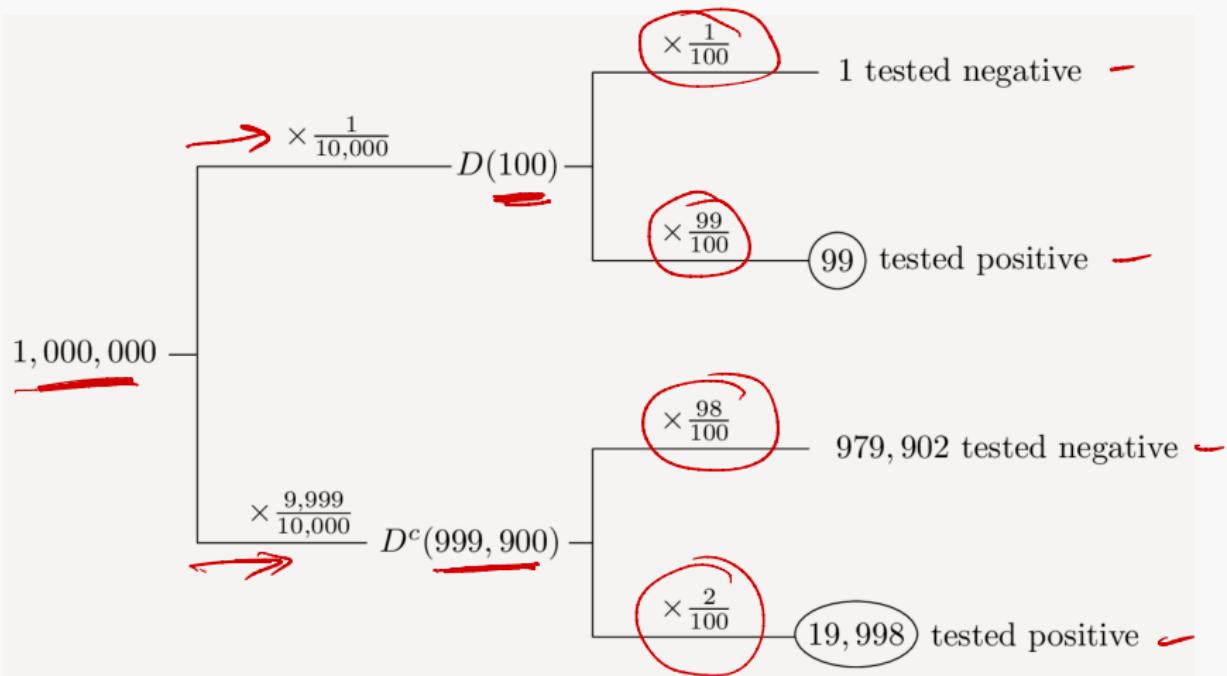


Fig.1.25 - Tree diagram for Example 1.26.

## Bayesian paradigm

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

Bayes' rule leads to *Bayesian statistics*.

- *Bayesian* interpretation: probability expresses a degree of belief in an event.  
Use *Bayes' rule* to update degree of belief based on observed data.
- *Frequentist* interpretation: probability is the long-run relative frequency of an event after many trials.
- Don't need to know for this course. More intuition here  
<https://www.youtube.com/watch?v=9wCnvr7Xw4E>

## Conditional probability

## Conditional independence

## Definition

Extend concept of *independence* to conditionally independent events.

### Conditional independence

Two events  $A$  and  $B$  are *conditionally independent* given an event  $C$  with  $P(C) > 0$  if

$$P(A \cap B|C) = P(A|C) P(B|C) \quad (7)$$

(Unconditional)  $\rightarrow$   $P(A \cap B) = P(A) P(B)$   
independence

## Example: Two coins

A box contains two coins: one regular coin and one two-headed coin ( $P(H) = 1$ ). Choose a coin at random and toss it twice. Define the following events.

- A: First coin-toss results in an H.
- B: Second coin-toss results in an H.
- C: Coin 1 (regular) has been selected.

*fair*

*I might expect*

$$P(B^c | A^c) > P(B^c | A)$$

Note that A and B are not independent, but they are *conditionally independent* given C. Find  $P(A|C)$ ,  $P(B|C)$ ,  $P(A \cap B|C)$ ,  $P(A)$ ,  $P(B)$ ,  $P(A \cap B)$ .

$$P(A|C) = 0.5 = P(B|C)$$

$$P(A \cap B|C) = \underbrace{P(A|C)}_{\text{cond ind}} \underbrace{P(B|C)}_{\text{cond ind}} = 0.25$$

$$P(A) = \underbrace{P(A|C)}_{0.5} \underbrace{P(C)}_{0.5} + \underbrace{P(A|C^c)}_{1} \underbrace{P(C^c)}_{1-0.5} = 0.75$$

$$P(B) = 0.75$$

$$P(A \cap B) = \underbrace{P(A \cap B|C)}_{0.25} \underbrace{P(C)}_{0.5} + \underbrace{P(A \cap B|C^c)}_{1} \underbrace{P(C^c)}_{0.5} = 0.625$$