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

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

1. Introduction
2. Conditional probability
   - Independence

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: Probabilty measure $P(\cdot)$

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

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

$$P\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = P(A_1) + P(A_2) + P(A_3) + \cdots \tag{1}$$

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.

Finding probabilities

Given a random experiment with a sample space Ω, 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:

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

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

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

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 Ω and arbitrary events $A, B \subset \Omega$, Definition 1 implies

1. $P(\varnothing) = 0$
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 $A \subset B$.

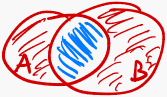
(Pictures for intuition; for formal proofs, see “Example 1.10” in §1.3.3 of textbook)

Finding probabilities: example

Suppose we have the following information:

1. There is a 60 percent chance that it will rain today. A = event of rain today
2. There is a 50 percent chance that it will rain tomorrow. B = event of rain tomorrow
3. There is a 30 percent chance that it does not rain either day. $(A \cup B)^c$

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.

a. $P(A \cup B) = 1 - P\left(\begin{matrix}\text{does not rain}\\ \text{either day}\end{matrix}\right) = 1 - 0.3 = 0.7$

b. $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.6 + 0.5 - 0.7 = 0.4$

c. $P(A \cap B^c) = P(A \setminus B) = P(A) - P(A \cap B) = 0.6 - 0.4 = 0.2$

d. $P(A \cup B) - P(A \cap B) = 0.7 - 0.4 = 0.3$

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 Ω is a countable set, this refers to a *discrete* probability model.

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

$$\boxed{A = \bigcup_{s \in A} \{s\}}$$

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

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

$\frac{1}{2^k}$

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?

$\sum_{k=1}^{\infty} \frac{1}{2^k} = 1$

k	1	2	3	4	5	...
winnings	-1	0	1	2	3	...
prob	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	

Let X = amount of winnings

- $P(1 \le X < 4) = P(\{X=1\} \cup \{X=2\} \cup \{X=3\})$
$= P(X=1) + P(X=2) + P(X=3)$
$= \frac{1}{8} + \frac{1}{16} + \frac{1}{32} = \boxed{\frac{7}{32}}$

- $P(X>1) = 1 - P(X \le 1) = 1 - [\frac{1}{2} + \frac{1}{4} + \frac{1}{8}] = \boxed{\frac{1}{8}}$

Probability models: discrete (equally likely outcomes)

Important special case: finite sample space Ω 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

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

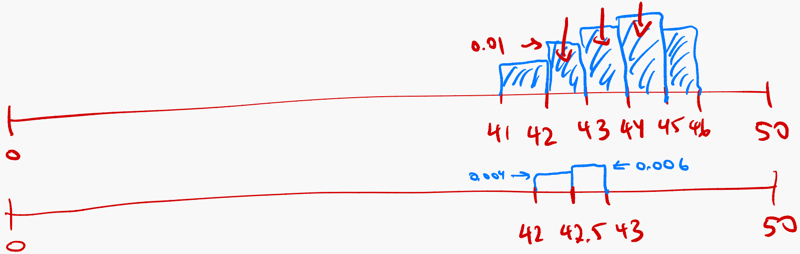
Probability models: continuous

Consider a sample space Ω 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 $T_{Ant} \in [42.5, 45)$?

If uniform assumption on finishing time,

then $P(T_{Ant} \in [42.5, 45)) = \dfrac{45 - 42.5}{50 - 0} = 0.05$



Conditional probability

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 $P(A)$.

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6}$$

- 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 $P(A|B)$.

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{|\{1,3\}|}{3} = \frac{2}{3}$$

Notation

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.

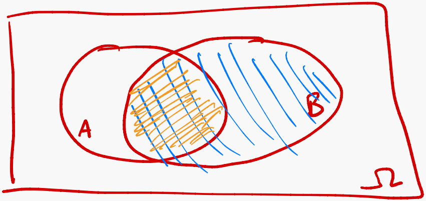
Definition 2: Conditional probability

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

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

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. $P(B|B) = 1$
- $P(A|B)$ 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)

If conditioning on event C, then slide 7 becomes

1. $P(\phi \mid C) = 0$
2. $P(A^c \mid C) = 1 - P(A \mid C)$
3. $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cap B \mid C)$

Important special cases

end of 9/29 lecture

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:

- When A is a subset of 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 $X_2 = 4$?

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:

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

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?

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.

- Guess that the result of the coin toss has nothing to do with tomorrow's weather.
- The probability of A should not depend on whether or not B happens.
- Two events are independent if one does not convey any info about the other.

Definition 1: Independent events

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

$$P(A \cap B) = P(A)P(B). \tag{3}$$

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

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

- “Independence means that conditional probability of one event given another is the same as the original (prior) probability.”
- Sometimes it is obvious if two given events are independent or not.
- Other times, we need need to check if they satisfy the independence condition (3).

Example

I pick a random number from $\{1, 2, 3, \cdots, 10\}$, 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?

Achtung

Independence does not mean disjoint! Disjoint does not mean independence!

$$\text{Independence} \neq \text{disjoint}$$

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