S&DS 241 Lecture 3

Conditional probability. Independence of events.

B-H: Sec 2.1-2.2, 2.4-2.5

Last time: Probability axioms

- Sample space: Ω
- Outcome: $\omega \in \Omega$
- Events: $A \subset \Omega$
- How to make new events: set operations

Probability axioms

Definition

1. Positivity: For any event A,
$$P(A) \geq 0,$$
2. Normalization:
$$P(\Omega) = 1$$
3. Additivity: Let A_i be mutually exclusive events, i.e., $A_i \cap A_j = \varnothing$ whenever $i \neq j$. Then
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Conditional probability: motivations

- The chance to attend office hour $\ll$ the chance to attend office hour if tomorrow is midterm

Conditional probability: motivations

- The chance to attend office hour $\ll$ the chance to attend office hour if tomorrow is midterm
- The chance of cancer among smokers $\gg$ the chance of lung cancer in general

Conditional probability: motivations

- The chance to attend office hour $\ll$ the chance to attend office hour if tomorrow is midterm
- The chance of cancer among smokers $\gg$ the chance of lung cancer in general
- What is the chance of blackjack given the cards dealt so far?
- ...

Conditional probability: definition

Definition

Given an event A with $P(A) > 0$, the conditional probability of event B given A is

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

Conditional probability: definition

Definition

Given an event A with $P(A) > 0$, the conditional probability of event B given A is

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

- Intuition: event changes the "probabilistic worldview"

$$\underbrace{P(B)}_{\text{before}} \xrightarrow{A \text{ occurred}} \underbrace{P(B|A)}_{\text{after}}$$

Conditional probability: definition

Definition

Given an event A with $P(A) > 0$, the conditional probability of event B given A is

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

- Intuition: event changes the "probabilistic worldview"

$$\underbrace{P(B)}_{\text{before}} \xrightarrow{A \text{ occurred}} \underbrace{P(B|A)}_{\text{after}}$$

- Statistical inference: conditioning on data

$$P(\text{drug is effective}|\text{result of the clinical trial})$$

Example: birthday

Consider the birthday of a stranger.

- $\Omega = \{\text{Janurary 1st}, \ldots, \text{December 31st}\}$

Example: birthday

Consider the birthday of a stranger.

- $\Omega = \{\text{Janurary 1st}, \ldots, \text{December 31st}\}$
- $P(\text{born on July 4th}) = \frac{1}{365}$

Example: birthday

Consider the birthday of a stranger.

- $\Omega = \{\text{Janurary 1st}, \ldots, \text{December 31st}\}$
- $P(\text{born on July 4th}) = \frac{1}{365}$
- $P(\text{born on July 4th}|\text{born on the 4th}) = \frac{1/365}{12/365} = \frac{1}{12}$

Example: birthday

Consider the birthday of a stranger.

- $\Omega = \{\text{Janurary 1st}, \ldots, \text{December 31st}\}$
- $P(\text{born on July 4th}) = \frac{1}{365}$
- $P(\text{born on July 4th}|\text{born on the 4th}) = \frac{1/365}{12/365} = \frac{1}{12}$
- $P(\text{born on July 4th}|\text{born in July}) = \frac{1/365}{31/365} = \frac{1}{31}$

Example: sampling without replacements

Consider an urn consisting of $\{{\color{blue}\bullet\bullet\bullet}{\color{red}\bullet\bullet}\}$. Draw one ball at random, remove it; then draw another at random.

Example: sampling without replacements

Consider an urn consisting of {●●●●●}. Draw one ball at random, remove it; then draw another at random.

- $A = \{$first draw is ●$\}$
- $B = \{$second draw is ●$\}$

Example: sampling without replacements

Consider an urn consisting of $\{{\color{blue}\bullet\bullet\bullet}{\color{red}\bullet\bullet}\}$. Draw one ball at random, remove it; then draw another at random.

- $A = \{\text{first draw is } {\color{blue}\bullet}\}$
- $B = \{\text{second draw is } {\color{blue}\bullet}\}$
- $P(B|A) =$

Example: sampling without replacements

Consider an urn consisting of {●●●●●}. Draw one ball at random, remove it; then draw another at random.

- $A = \{$first draw is ●$\}$
- $B = \{$second draw is ●$\}$
- $P(B|A) = \frac{1}{2}$, by intuition

Example: sampling without replacements

Consider an urn consisting of {●●●●●}. Draw one ball at random, remove it; then draw another at random.

- $A = \{$first draw is ●$\}$
- $B = \{$second draw is ●$\}$
- $P(B|A) = \frac{1}{2}$, by intuition
- Verify this by definition:

Example: sampling without replacements

Consider an urn consisting of {●●●●●}. Draw one ball at random, remove it; then draw another at random.

- $A = \{$first draw is ●$\}$
- $B = \{$second draw is ●$\}$
- $P(B|A) = \frac{1}{2}$, by intuition
- Verify this by definition:
  - $P(A) = \frac{3}{5}$

Example: sampling without replacements

Consider an urn consisting of $\{\color{blue}\bullet\bullet\bullet\color{red}\bullet\bullet\}$. Draw one ball at random, remove it; then draw another at random.

- $A = \{\text{first draw is } \color{blue}\bullet\}$
- $B = \{\text{second draw is } \color{blue}\bullet\}$
- $P(B|A) = \frac{1}{2}$, by intuition
- Verify this by definition:
  - $P(A) = \frac{3}{5}$
  - $P(A \cap B) = P(\text{both are } \color{blue}\bullet\color{black}) = \frac{\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10}$

Example: sampling without replacements

Consider an urn consisting of {●●●●●}. Draw one ball at random, remove it; then draw another at random.

- $A = \{$first draw is ●$\}$
- $B = \{$second draw is ●$\}$
- $P(B|A) = \frac{1}{2}$, by intuition
- Verify this by definition:
  - $P(A) = \frac{3}{5}$
  - $P(A \cap B) = P(\text{both are ●}) = \frac{\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10}$
  - $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1}{2}$

Axioms are inherited

Conditional probability satisfies three axioms on original sample space Ω:

1. Positivity: $B \subset \Omega$

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

Axioms are inherited

Conditional probability satisfies three axioms on original sample space Ω:

1. Positivity: $B \subset \Omega$

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

2. Normalization:

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

Axioms are inherited

Conditional probability satisfies three axioms on original sample space Ω:

1. Positivity: $B \subset \Omega$

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

2. Normalization:

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

3. Additivity: Let $B_i \subset \Omega$ be mutually exclusive. Then

$$P\left(\bigcup_{i=1}^{\infty} B_i \middle| A\right) \stackrel{\text{exercise}}{=} \sum_{i=1}^{\infty} P(B_i|A)$$

Axioms are inherited

Alternatively, conditional probability satisfies three axioms on the new sample space A (the outcomes outside A are now ruled out)

1. Positivity: $B \subset A$,

$$P\left(B|A\right) = \frac{P(B)}{P(A)} \geq 0$$

Axioms are inherited

Alternatively, conditional probability satisfies three axioms on the new sample space A (the outcomes outside A are now ruled out)

1. Positivity: $B \subset A$,

$$P\left(B|A\right) = \frac{P(B)}{P(A)} \geq 0$$

2. Normalization:

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

Axioms are inherited

Alternatively, conditional probability satisfies three axioms on the new sample space A (the outcomes outside A are now ruled out)

1. Positivity: $B \subset A$,

$$P\left(B|A\right) = \frac{P(B)}{P(A)} \geq 0$$

2. Normalization:

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

3. Additivity: Let $B_i \subset A$ be mutually exclusive. Then

$$P\left(\bigcup_{i=1}^{\infty} B_i \middle| A\right) = \sum_{i=1}^{\infty} P\left(B_i|A\right)$$

Consequences

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

Consequences

- $P(B^c|A) = 1 - P(B|A)$
- A and B mutually exclusive $\implies P(B|A) = 0$
  e.g., $P(\text{born on July 4th}|\text{born in March}) = 0$

Consequences

- $P(B^c|A) = 1 - P(B|A)$
- A and B mutually exclusive $\implies P(B|A) = 0$
  e.g., $P(\text{born on July 4th}|\text{born in March}) = 0$
- Equiprobable experiment remains equiprobable after conditioning
  - $P(A) = \frac{|A|}{|\Omega|}$
  - $P(A \cap B) = \frac{|A \cap B|}{|\Omega|}$
  - $P(B|A) = \frac{|A \cap B|}{|A|}$.

Example: coin experiment

- Three coins in my pocket
  1. one with head on both sides
  2. one with tail on both sides
  3. one with head and tail.
- Pick one at random and flip it

Example: coin experiment

- Three coins in my pocket
  1. one with head on both sides
  2. one with tail on both sides
  3. one with head and tail.
- Pick one at random and flip it
- $P(\text{front is H}) = ?$

Example: coin experiment

- Three coins in my pocket
  1. one with head on both sides
  2. one with tail on both sides
  3. one with head and tail.
- Pick one at random and flip it
- $P(\text{front is H}) = ?$
- $P(\text{back is also H} | \text{front is H}) = ?$

Example: coin experiment

- Three coins in my pocket
  1. one with head on both sides
  2. one with tail on both sides
  3. one with head and tail.
- Pick one at random and flip it

Example: coin experiment

- Three coins in my pocket
  1. one with head on both sides
  2. one with tail on both sides
  3. one with head and tail.
- Pick one at random and flip it
- Sample space: possibilities of front

coin 1	coin 2	coin 3
H_1	T_1	H_3
H_2	T_2	T_3

Example: coin experiment

- Three coins in my pocket
  1. one with head on both sides
  2. one with tail on both sides
  3. one with head and tail.
- Pick one at random and flip it
- Sample space: possibilities of front

coin 1	coin 2	coin 3
H_1	T_1	H_3
H_2	T_2	T_3

- $P(\text{front is H}) = \frac{3}{6} = \frac{1}{2}$;

Example: coin experiment

- Three coins in my pocket
  1. one with head on both sides
  2. one with tail on both sides
  3. one with head and tail.
- Pick one at random and flip it
- Sample space: possibilities of front

coin 1	coin 2	coin 3
H_1	T_1	H_3
H_2	T_2	T_3

- $P(\text{front is H}) = \frac{3}{6} = \frac{1}{2}$; similarly, $P(\text{back is H}) = \frac{1}{2}$

Example: coin experiment

- Three coins in my pocket
  1. one with head on both sides
  2. one with tail on both sides
  3. one with head and tail.
- Pick one at random and flip it
- Sample space: possibilities of front

coin 1	coin 2	coin 3
H_1	T_1	H_3
H_2	T_2	T_3

- $P(\text{front is H}) = \frac{3}{6} = \frac{1}{2}$; similarly, $P(\text{back is H}) = \frac{1}{2}$
- $P(\text{both sides are H}) = P(\text{coin 1 is chosen}) = \frac{1}{3}$

Example: coin experiment

- Three coins in my pocket
  1. one with head on both sides
  2. one with tail on both sides
  3. one with head and tail.
- Pick one at random and flip it
- Sample space: possibilities of front

coin 1	coin 2	coin 3
H_1	T_1	H_3
H_2	T_2	T_3

- $P(\text{front is H}) = \frac{3}{6} = \frac{1}{2}$; similarly, $P(\text{back is H}) = \frac{1}{2}$
- $P(\text{both sides are H}) = P(\text{coin 1 is chosen}) = \frac{1}{3}$
- $P(\text{back is H}|\text{front is H}) = \frac{1/3}{1/2} = \frac{2}{3} > P(\text{back is H}) = \frac{1}{2}$: Why higher?

Example: sampling **without** replacements

Consider an urn consisting of {●●●●●}. Draw one ball at random, remove it; then draw another at random.

- $A = \{$first draw is ●$\}$
- $B = \{$second draw is ●$\}$
- $P(B|A) = \frac{1}{2} < P(B) \overset{\text{why?}}{=} \frac{3}{5}$: Why lower?

Example: sampling **with** replacements

Consider an urn consisting of $\{$●●●●●$\}$. Draw one ball at random, put it back; then draw another at random.

- $A = \{$first draw is ●$\}$
- $B = \{$second draw is ●$\}$
- $P(B|A) = P(B) = \frac{3}{5}$: Why equal?

So we have witnessed...

- All of the following are possible
  - $P(B|A) > P(B)$
  - $P(B|A) < P(B)$
  - $P(B|A) = P(B)$

So we have witnessed...

- All of the following are possible
  - $P(B|A) > P(B)$
  - $P(B|A) < P(B)$
  - $P(B|A) = P(B)$
- "Independence" $\Leftrightarrow$ conditioning has no effect

Independence of two events

Definition (Independence)

Given events A, B with $P(A) > 0, P(B) > 0$, we say A and B are independent if

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

Independence of two events

Definition (Independence)

Given events A, B with $P(A) > 0, P(B) > 0$, we say A and B are independent if

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

Remarks

- Independence is a symmetric notion:

$$\begin{aligned} & P(A|B) = P(A) \\ \Leftrightarrow\ & P(B|A) = P(B) \\ \Leftrightarrow\ & P(A \cap B) = P(A)P(B) \end{aligned}$$

Simple examples of independent events

- Flip two coins: {first toss is H} and {second toss is T} — implicitly assumed to be "physically independent"

Simple examples of independent events

- Flip two coins: {first toss is H} and {second toss is T} — implicitly assumed to be "physically independent"
- Throw a die: {even} and {divisible by 3} (Why?)

Observations

- A and B independent $\implies$ A^c and B independent, A^c and B^c independent, ...

Observations

- A and B independent $\implies$ A^c and B independent, A^c and B^c independent, ...
- A and A^c

Observations

- A and B independent $\implies$ A^c and B independent, A^c and B^c independent, ...
- A and A^c are NOT independent (in fact highly dependent!): $P(A|A^c) = 0$
- $A \subset B$

Observations

- A and B independent $\implies$ A^c and B independent, A^c and B^c independent, ...
- A and A^c are NOT independent (in fact highly dependent!): $P(A|A^c) = 0$
- $A \subset B$ are NOT independent (in fact highly dependent!): $P(B|A) = 1$

Example

- A statistics professor had to take a flight for the first time and was terrified of the prospect that someone will bring a knife on the plane.

Example

- A statistics professor had to take a flight for the first time and was terrified of the prospect that someone will bring a knife on the plane.
- He learned that <u>one out of a thousand</u> people does that.

Example

- A statistics professor had to take a flight for the first time and was terrified of the prospect that someone will bring a knife on the plane.
- He learned that one out of a thousand people does that.
- Using his probabilistic expertise, he decided to bring a knife himself because the chance that two people carry knives is one in a million, which is much more reassuring.

Example

- A statistics professor had to take a flight for the first time and was terrified of the prospect that someone will bring a knife on the plane.
- He learned that one out of a thousand people does that.
- Using his probabilistic expertise, he decided to bring a knife himself because the chance that two people carry knives is one in a million, which is much more reassuring.
- Do you agree with his reasoning?

Independence of three events

Definition

Events A, B and C are mutually independent if

$$\left.\begin{aligned} P(AB) &= P(A)P(B) \\ P(BC) &= P(B)P(C) \\ P(CA) &= P(C)P(A) \end{aligned}\right\} \text{ ``pairwise independence''}$$

and

$$P(ABC) = P(A)P(B)P(C)$$

Shorthand: AB stands for $A \cap B$, ABC for $A \cap B \cap C$, etc.

Example

$$A = \{\text{Alice and Bob are born on the same day}\}$$
$$B = \{\text{Bob and Charlie are born on the same day}\}$$
$$C = \{\text{Charlie and Alice are born on the same day}\}$$

Are they mutually independent?

Example

$$A = \{\text{Alice and Bob are born on the same day}\}$$
$$B = \{\text{Bob and Charlie are born on the same day}\}$$
$$C = \{\text{Charlie and Alice are born on the same day}\}$$

Are they mutually independent?

- $P(A) = P(B) = P(C) = \frac{1}{365}$

Example

$$A = \{\text{Alice and Bob are born on the same day}\}$$

$$B = \{\text{Bob and Charlie are born on the same day}\}$$

$$C = \{\text{Charlie and Alice are born on the same day}\}$$

Are they mutually independent?

- $P(A) = P(B) = P(C) = \frac{1}{365}$
- A, B, C are pairwise independent:

$$P(AB) = P(\text{all three have same birthday}) = \frac{1}{365^2} = P(A)P(B)$$

Example

$$A = \{\text{Alice and Bob are born on the same day}\}$$
$$B = \{\text{Bob and Charlie are born on the same day}\}$$
$$C = \{\text{Charlie and Alice are born on the same day}\}$$

Are they mutually independent?

- $P(A) = P(B) = P(C) = \frac{1}{365}$
- A, B, C are pairwise independent:

$$P(AB) = P(\text{all three have same birthday}) = \frac{1}{365^2} = P(A)P(B)$$

- A, B, C are dependent: $AB = ABC$ so $P(C|AB) = 1$ and

$$P(ABC) = \frac{1}{365^2} \neq P(A)P(B)P(C)$$

Example

$$A = \{\text{Alice and Bob are born on the same day}\}$$
$$B = \{\text{Bob and Charlie are born on the same day}\}$$
$$C = \{\text{Charlie and Alice are born on the same day}\}$$

Are they mutually independent?

- $P(A) = P(B) = P(C) = \frac{1}{365}$
- A, B, C are pairwise independent:

$$P(AB) = P(\text{all three have same birthday}) = \frac{1}{365^2} = P(A)P(B)$$

- A, B, C are dependent: $AB = ABC$ so $P(C|AB) = 1$ and

$$P(ABC) = \frac{1}{365^2} \neq P(A)P(B)P(C)$$

- Lesson: dependence is revealed when examined jointly

Consequences

$$
\begin{aligned}
& A, B, C \text{ are mutually independent} \\
\implies & A \text{ and } B \cup C \text{ independent} \\
\implies & A \text{ and } B^c \cap C \text{ independent, etc}
\end{aligned}
$$

Knowing that A, B, C are pairwise independent is not enough for these implications.

Independence of n events

Definition: check independence of all pairs, all triples, all 4-tuples, ...

Formally:

Definition

Events $A_1, A_2, \ldots, A_n$ are mutually independent if

$$P(A_{i_1} A_{i_2} \ldots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \ldots P(A_{i_k}), \quad \text{“}k\text{-way independence”}$$

for all $k \geq 2$ and all $1 \leq i_1 < i_2 < \ldots i_k \leq n$

Example: backups

Example

A data center consists of n hard drives storing the same file. Suppose each of them fails with probability 40% independently. How many hard drives are needed to boost the reliability to at least 99.9%?

Example: backups

Example

A data center consists of n hard drives storing the same file. Suppose each of them fails with probability 40% independently. How many hard drives are needed to boost the reliability to at least 99.9%?

- $A_i = \{i\text{th hard drive fails}\}$
- Independence $\implies P(A_1 \cap A_2 \cap \cdots A_n) = 0.4^n$
- $P(\text{file accessible}) = 1 - P(A_1 \cap A_2 \cap \cdots A_n) = 1 - 0.4^n \geq 0.999 \implies n \geq 8$