ECE 203 – Section 2
Independence and conditioning

- Independence
- Conditional probabilities
- Law of total probability
- Bayes' formula, prior and posterior probabilities
- Bayesian classification

The slides have been prepared based on the lecture notes of Prof. Patrick Mitran.

- Conditional probability is one of the most important concepts in this course.
- It lets us compute/update probabilities when partial information is revealed. It is also a tool to compute probabilities.
- *Example:* Suppose we toss two dice. There are 36 outcomes:

$$\begin{aligned} S = \{&(1,1),(1,2),\ldots,(1,6),\\ &(2,1),(2,2),\ldots,(2,6),\\ &\quad\vdots \qquad\qquad\quad \vdots\\ &(6,1),(6,2),\ldots,(6,6)\} \end{aligned}$$

- What is the probability that the sum is 8?
- This event is $E = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$. So $P[E] = 5/36$.

- Now, assume we roll the 1st die, but not the 2nd, and we get 3.
- What is the probability that the sum will be 8?
- All possible outcomes given this new information are

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

- The other 30 cases are inconsistent with the 1st die roll. *They now have zero probability.*

- The 6 cases in F had the same probability before the 1st die was rolled.
- So they should now be equally likely after we know the outcome of 1st die roll, i.e., each has probability $1/6$.
- After the 1st die roll was revealed, i.e., after F was revealed to have occurred:

$$\{\text{sum } = 8\} = EF = \{(3,5)\}$$

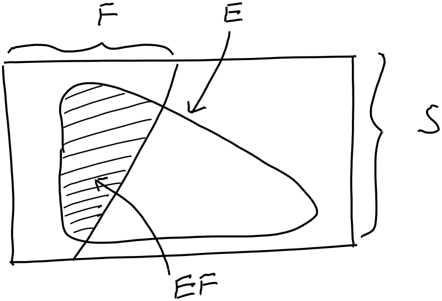
and this has probability $1/6$.

We say that *the probability of E given F has occurred* is $1/6$, or

$$P[E \mid F] = 1/6$$

- For brevity, we often *colloquially* say that $P(E \mid F)$ is a *conditional probability of E given F*.

- Let's generalize. In particular, let's not assume the elements of S are equally likely.



- If F has occurred, then for E to occur, EF must occur.
- If F has occurred, our sample space S is reduced to F.
- So if F has occurred, probabilities should be computed relative to F:

Conditional Probability (cont.)

Definition: If $P[F] > 0$, then

$$P[E \mid F] = \frac{P[EF]}{P[F]}$$

- *Example:* A coin is flipped twice. What is the probability of two heads if
  1. first flip is heads?
  2. at least one flip is heads?

- In the first case, $S = \{hh, ht, th, tt\}, E = \{hh\}$ and $F = \{ht, hh\}$. So

$$P[E \mid F] = \frac{P[EF]}{P[F]} = \frac{P[hh]}{P[F]} = \frac{1/4}{1/2} = 1/2$$

- In the second case, $F = \{hh, ht, th\}$. So

$$P[E \mid F] = \frac{P[EF]}{P[F]} = \frac{1/4}{3/4} = 1/3$$

More examples

- Two dice are rolled. Let

$$E = \{\text{ max of both rolls is } 4\}$$

and

$$F = \{\text{ min of both rolls is } 3\}$$

- What is $P[E \mid F]$?
- To solve this, we notice

$$E = \{(1,4),(2,4),(3,4),(4,4),(4,3),(4,2),(4,1)\}$$

and

$$F = \{(6,3),(5,3),(4,3),(3,3),(3,4),(3,5),(3,6)\}$$

- So,

$$P[E \mid F] = \frac{P[EF]}{P[F]} = \frac{2/36}{7/36} = 2/7$$

Example: Baseball and hockey teams

- A city has a baseball and a hockey team. Both teams are playing on the same day.
- The baseball and hockey teams each have probability 1/2 and 1/3 of winning their games.
- The probability of at least one team wining is 3/5.
- *If only one team wins its game, what is the probability that it was the baseball team?*

- Let's denote

$$WW = \{\text{both win}\}$$
$$LL = \{\text{both lose}\}$$
$$WL = \{\text{baseball wins, hockey loses}\}$$
$$LW = \{\text{baseball loses, hockey wins}\}$$

- Then we are given that

$$P[WW] + P[WL] = 1/2$$
$$P[WW] + P[LW] = 1/3$$
$$P[WW] + P[WL] + P[LW] = 3/5$$

- Consequently, we have

$$P[LL] = 2/5$$
$$P[WL] = 3/5 - 1/3 = 4/15$$
$$P[LW] = 3/5 - 1/2 = 1/10$$
$$P[WW] = 1 - 2/5 - 4/15 - 1/10 = 7/30$$

- Finally,

$$P\,[\text{baseball team wins} \mid \text{one of the teams wins}] =$$

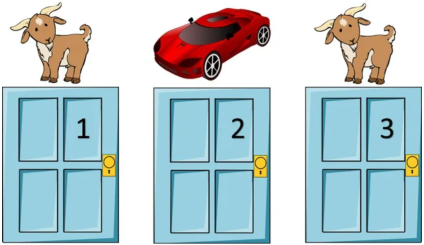
$$= P\,[\{WL, WW\} \mid \{WL, LW\}] =$$

$$= \frac{P[\{WL, WW\} \cap \{WL, LW\}]}{P[\{WL, LW\}]} =$$

$$= \frac{P[\{WL\}]}{P[\{WL, LW\}]} = \frac{4/15}{4/15 + 1/10} = 8/11$$

You are at a TV show; You are given the opportunity to select one closed door of three, behind one of which there is a prize. The other two doors hide goats. Once you have made your selection, the TV host will open one of the remaining doors, revealing that it does not contain the prize. He then asks you if you would like to switch your selection to the other unopened door, or stay with your original choice.

Do you switch ?



Important: For $P[F] > 0$, $P[\cdot \mid F]$ satisfies the same axioms as $P[\cdot]$.

[A1]

$$P[E \mid F] = P[EF]/P[F] \geq 0 \qquad (\text{since } P[EF] \geq 0)$$
$$P[E \mid F] = P[EF]/P[F] \leq 1 \qquad (\text{since } EF \subset F)$$

[A2]

$$P[S \mid F] = P[SF]/P[F] = P[F]/P[F] = 1$$

[A3] Let $E_1 \cap E_2 = \emptyset$. Then $E_1 F \cap E_2 F = \emptyset$.

$$P[E_1 \cup E_2 \mid F] = P[(E_1 \cup E_2)F]/P[F] =$$
$$= P[E_1 F \cup E_2 F]/P[F] = P[E_1 F]/P[F] + P[E_2 F]/P[F] =$$
$$= P[E_1 \mid F] + P[E_2 \mid F]$$

Multiplication Rule

- Since F_1F_2 is a set, we also write $P[E|F_1F_2] = P[EF_1F_2]/P[F_1F_2]$, etc.

- Now we have:

$$P[E_1E_2\cdots E_n] = P[E_1]\cdot\frac{P[E_1E_2]}{P[E_1]}\cdot\frac{P[E_1E_2E_3]}{P[E_1E_2]}\cdot\cdots$$
$$\cdots\cdot\frac{P[E_1E_2\dots E_{n-1}]}{P[E_1E_2\dots E_{n-2}]}\cdot\frac{P[E_1E_2\dots E_n]}{P[E_1E_2\dots E_{n-1}]}$$

or, equivalently,

$$\boxed{\begin{gathered}P[E_1E_2\cdots E_n] = P[E_1]\cdot P[E_2 \mid E_1]\cdot P[E_3 \mid E_1E_2]\cdot\cdots\\ \cdots\cdot P[E_n \mid E_1E_2\cdots E_{n-1}]\end{gathered}}$$

which is known as *multiplication rule*.

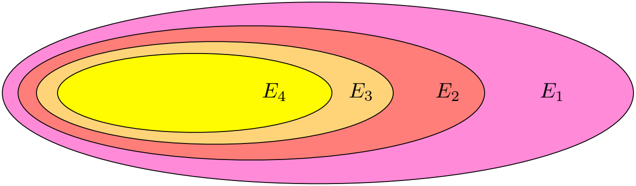
Example

- 4 graduate and 16 undergraduate students are randomly divided into 4 groups of 5.
- What is the probability that each group has exactly one graduate student?
- To solve this problem, we first denote:

$$\begin{aligned} \text{Let} \quad E_1 &= \{\text{grad \#1 is in a group}\}, \\ E_2 &= \{\text{grad \#1 and \#2 are in different groups}\} \\ E_3 &= \{\text{grad \#1, \#2 and \#3 are in different groups}\} \\ E_4 &= \{\text{all grads are in different groups}\} \end{aligned}$$

Example (cont.)

$$E_4 \subset E_3 \subset E_2 \subset E_1$$



Then

$$\begin{aligned} P[E_4] &= P[E_1 E_2 E_3 E_4] \\ &= P[E_1] P[E_2|E_1] P[E_3|E_1 E_2] P[E_4|E_1 E_2 E_3] \end{aligned}$$

with

$$\begin{aligned} &P[E_1] = 1 \\ &P[E_2|E_1] = 15/19 \\ &P[E_3|E_1 E_2] = 10/18 \\ &P[E_4|E_1 E_2 E_3] = 5/17 \end{aligned}$$

Law of Total Probability (Special Case): Let $E, F \subset S$. Then,

$$E = ES = E(F \cup F^c) = EF \cup EF^c$$

and, consequently,

$$P[E] = P[EF] + P[EF^c] = P[E|F]P[F] + P[E|F^c]P[F^c]$$

Example

- You enter a tournament in which there are two types of players, *viz.* Type I and Type II.
- There are 30% and 70% of Type I and Type II players, respectively, implying that your chance to play against a Type I (resp., Type II) opponent is 0.3 (resp., 0.7).
- Your chance to win against a Type I opponent is 0.4.
- Your chance to win against a Type II opponent is 0.6.
- *If you play against a random opponent of unknown type, what is the probability of you winning $P(you\ win)$?*

- Let introduce some notations first.

$$E = \{\text{you win}\}$$

$$F = \{\text{your opponent is of Type I}\}$$

and, therefore,

$$F^c = \{\text{your opponent is of Type II}\}$$

- Consequently,

$$P(E) = P(E \mid F)\, P(F) + P(E \mid F^c)\, P(F^c) =$$

$$= 0.4 \cdot 0.3 + 0.6 \cdot 0.7 = 0.54.$$

Law of Total Probability (General Case): Let $F_1, F_2, \ldots, F_n$ partition S. Then,

$$E = ES = E\left(\bigcup_{i=1}^{n} F_i\right) = \bigcup_{i=1}^{n} (EF_i)$$

and, consequently,

$$P[E] = P\left[\cup_{i=1}^{n} (EF_i)\right] = \sum_{i=1}^{n} P[EF_i] = \sum_{i=1}^{n} P[E|F_i] P[F_i]$$

Example

- You flip a coin 3 times. If the number of heads is less than 2, you flip the coin two more times. What is the probability of at least 3 heads?

- To solve this problem, we introduce the following notations (with $\#H$ standing for "number of heads").

$$F_i = \{\#H \text{ in first 3 flips } = i\}$$

with $i = 1, 2, 3$ and

$$E = \{\#H \geq 3\}$$

- We have

$$P[F_0] = 1/8$$
$$P[F_1] = 3/8$$
$$P[F_2] = 3/8$$
$$P[F_3] = 1/8$$

Example (cont.)

- We also notice that

$$
\begin{aligned}
P[E \mid F_0] &= 0 \\
P[E \mid F_1] &= P\,[\text{both extra flips are heads}] = 1/4 \\
P[E \mid F_2] &= 0 \\
P[E \mid F_3] &= 1
\end{aligned}
$$

- Consequently,

$$P(E) = P[E|F_0]P[F_0] + P[E|F_1]P[F_1] + P[E|F_2]P[F_2] + P[E|F_3]P[F_3] =$$

$$= 0 \cdot 1/8 + 1/4 \cdot 3/8 + 0 \cdot 3/8 + 1 \cdot 1/8 = 7/32.$$

- Let $F_1, F_2, \ldots, F_n$ partition S.
- Suppose we know $P(E \mid F_i)$ for $i = 1, 2, \ldots, n$.
- Then, for any $1 \le j \le n$, we have the following result.

Bayes' Theorem

$$P[F_j \mid E] = \frac{P[EF_j]}{P[E]} = \frac{P[E|F_j]P[F_j]}{\sum_{i=1}^{n} P[E \mid F_i]P[F_i]} =$$

$$= \frac{P[E \mid F_j]P[F_j]}{P[E \mid F_1]P[F_1] + P[E \mid F_2]P[F_2] + \cdots + P[E \mid F_n]P[F_n]}$$

- One can say that, in this way, Bayes' Theorem allows "reversing the order of conditioning".

- Before any partial information is revealed (i.e., observing E occurs), the probabilities are:

$$P[F_1], P[F_2], \ldots, P[F_n]$$

- They are known as *"prior probabilities"*.

- After observing E occur, they are revised as:

$$P[F_1 \mid E], P[F_2 \mid E], \ldots, P[F_n \mid E]$$

- The new probabilities are called *"posterior probabilities"*.

- Posterior probabilities are the driving force behind practical inference (e.g., classification, pattern recognition, detection, etc.)

Example

- A 3-card deck has
  1. one card with red on both sides
  2. one card with black on both sides
  3. one card with red on one side + black on the other

- One side of 1 card is picked at random. It is red.

- *What is the probability that the other side is black?*

- To solve this problem, we define:

$$S = \{RR, RB, BB\} \quad \text{and} \quad R = \{\text{side shown is red}\}$$

- Thus, we have

$$P[RB \mid R] = \frac{P[RB \cap R]}{P[R]} =$$

$$= \frac{P[R \mid RB]P[RB]}{P[R \mid RR]P[RR] + P[R \mid RB]P[RB] + P[R \mid BB]P[BB]} =$$

$$= \frac{\frac{1}{2} \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = 1/3$$

- Note that, if you see red, there are 3 different ways this could happen (one side of RB + two sides of RR). Yet, only 1 results in black on the other side.

Another example

- A person is missing, but is known to be in 1 of 3 locations (each with probability 1/3).
- If the person is in location i, searching there has probability $1 - \beta_i$ of finding them (where β_j is known as *overlook probability*).
- *What is probability that the person is in location j if search of location 1 is unsuccessful?*
- To solve this problem, we define:

$$R_i = \{\text{the person is in location } i\}$$

$$E = \{\text{search of location 1 unsuccessful}\}$$

Another example (cont.)

- Then, we have

$$P[R_1|E] = \frac{P[R_1E]}{P[E]} = \frac{P[E|R_1]P[R_1]}{P[E|R_1]P[R_1] + P[E|R_2]P[R_2] + P[E|R_3]P[R_3]} =$$

$$= \frac{\beta_1 \cdot \frac{1}{3}}{\beta_1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{\beta_1}{2 + \beta_1}$$

- Similarly,

$$P[R_2|E] = \frac{P[R_2E]}{P[E]} = \frac{P[E|R_2]P[R_2]}{P[E|R_1]P[R_1] + P[E|R_2]P[R_2] + P[E|R_3]P[R_3]} =$$

$$= \frac{1 \cdot \frac{1}{3}}{\beta_1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{1}{2 + \beta_1}$$

- Finally, $P[R_3|E]$ is same as $P[R_2|E]$.

Yet another example

- A blood test is 95% effective in detecting a disease when it is present.
- It has a 1% false positive rate when it is not present and 0.5% of people have the disease.
- **Question (a):** *If a random person tests positive, what is the probability that the disease is present?*
- **Question (b):** *If a random person tests negative, what is the probability that the disease is present?*
- To solve this problem, we define:

$$E = \{\text{positive result}\} \quad \text{and} \quad F = \{\text{disease present}\}$$

Yet another example (cont.)

- Thus, for Question (a) we have:

$$P[F|E] = \frac{P[EF]}{P[E]} = \frac{P[E|F]P[F]}{P[E|F]P[F] + P[E|F^c]P[F^c]} =$$
$$= \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} \approx 0.323$$

- Similarly, for Question (b) we have:

$$P[F|E^c] = \frac{P[E^cF]}{P[E^c]} = \frac{P[E^c|F]P[F]}{P[E^c|F]P[F] + P[E^c|F^c]P[F^c]}$$
$$= \frac{0.05 \times 0.005}{0.05 \times 0.005 + 0.99 \times 0.995} \approx 0.000254$$

Independent Events

Two events E and F are said to be **independent** if

$$P[EF] = P[E]\,P[F]$$

Two events that are not independent are said to be **dependent**.

- From previous examples, $P[E \mid F]$ is not necessarily the same as $P[E]$.
- But, if E and F are independent (and $P[F] > 0$), then

$$P[E|F] = \frac{P[EF]}{P[F]} = \frac{P[E]P[F]}{P[F]} = P[E]$$

Thus, the independence of E and F always suggests

$$P[E \mid F] = P[E]$$

- **Problem 1:** A card is randomly selected from a standard deck.
  - Let $E = \{\text{card is a 1}\}$ and $F = \{\text{card is a spade}\}$.
  - Then $P[EF] = 1/52$, $P[E] = 4/52$, and $P[F] = 13/52$.
  - Thus $P[EF] = P[E]P[F]$, implying E and F are independent.

- **Problem 2:** Suppose $EF = \emptyset$, with $P[E] > 0$ and $P[F] > 0$. Are E and F independent?
  - On one hand, we have
  $$P[E|F] = \frac{P[EF]}{P[F]} = \frac{P[\emptyset]}{P[F]} = \frac{0}{P[F]} = 0$$
  - On the other hand, we know that $P[E] > 0$. So, no.

- **Problem 3:** Two 6-sided dice are rolled.
  - Let $E_1 = \{\text{sum is 6}\}$, $E_2 = \{\text{sum is 7}\}$, $F = \{\text{first die is 4}\}$, and $G = \{\text{second die is 3}\}$.
  - Then $P[E_1F] = P[(4,2)] = 1/36$, while $P[E_1]P[F] = 5/36 \cdot 1/6 \neq 1/36$.
  - Also, $P[E_2F] = P[(4,3)] = 1/36$ and $P[E_2]P[F] = 1/6 \cdot 1/6 = 1/36$.
  - So E_1 and F are not independent, but E_2 and F are independent.
  - Similarly, E_2 and G are independent.

Theorem

If E and F are independent, then E and F^c are independent

Proof.

$$P[E] = P[EF \cup EF^c] = P[EF] + P[EF^c] = P[E]P[F] + P[EF^c]$$

Therefore,

$$P[EF^c] = P[E] - P[E]P[F] = P[E](1 - P[F]) = P[E]P[F^c]$$

$\square$

- **Question:** If E is independent of both F and G, is E independent of FG?
- **Answer:** Not necessarily.

  In the example with two dice, E_2 is independent of both F and G. Yet, $P[E_2|FG] = P[\{\text{sum is } 7\}|(4,3)] = 1$, while $P[E_2] = 6/36$.

Conditional Independence

Events E and F are called *conditionally independent given G*, when

$$P[EF|G] = P[E|G]P[F|G]$$

- What does this mean?
- First, we note that

$$P[E|G]P[F|G] = P[EF|G] = \frac{P[EFG]}{P[G]} = \frac{P[E|FG] \cdot P[F|G] \cdot P[G]}{P[G]}$$

- So, this is equivalent to $P[E|FG] = P[E|G]$.
- **In words:** If G is known to have occurred, the additional information that F occurred does not change the probability of E.

Conditional Independence (cont.)

The 3 events E, F and G are said to be *independent* if

$$P[EFG] = P[E]P[F]P[G]$$
$$P[EF] = P[E]P[F]$$
$$P[EG] = P[E]P[G]$$
$$P[FG] = P[F]P[G]$$

In this case, E is independent of any event formed from F and G.

- **Example 1:** $P[E(FG)] = P[EFG] = P[E]P[F]P[G] = P[E]P[FG]$.
- **Example 2:**

$$\begin{aligned} &P[E(F \cup G)] = P[EF \cup EG] = P[EF] + P[EG] - P[EF \cap EG] = \\ &= P[E]P[F] + P[E]P[G] - P[E]P[FG] = \\ &= P[E](P[F] + P[G] - P[FG]) = \\ &= P[E]P[F \cup G] \end{aligned}$$

Independence

Events $E_1, E_2, ...E_n$ are said to be independent if

$$P\left[\bigcap_{i \in A} E_i\right] = \prod_{i \in A} P[E_i]$$

for every $A \subset \{1, ..., n\}$.

An infinite set of events $E_1, E_2, \ldots$ are independent, if every finite subset of the events are independent.

Example

- A system is composed of n components. Each component functions/fails independently of any other component.
- Component i has probability p_i of functioning. If at least one component functions, the system functions.
- *What is the probability that the system functions?*
- To solve this problem, let $A_i = \{i\text{-th component functions}\}$. Then

$$
\begin{aligned}
P[\text{system functions}] &= 1 - P[\text{system does not function}] = \\
&= 1 - P[\text{all components fail}] = \\
&= 1 - P[\cap_i A_i^c] = \\
&= 1 - \underbrace{P[A_1^c]P[A_2^c]\cdots P[A_n^c]}_{\textit{by independence}} = \\
&= 1 - (1-p_1)(1-p_2)\cdots(1-p_n).
\end{aligned}
$$

- Sometimes each E_i is the outcome of one instance of a sequence of repeated sub-experiments, e.g., $E_i = \{i\text{-th coin toss is heads}\}$.
- These sub-experiments are often called *trials* (or *repeated trials*).

Example

- Independent trials that consist of rolling a pair of fair dice are performed.
- The outcome of a roll is the sum of the dice.
- *What is the probability that an outcome of 5 appears before an outcome of 7?*
- To solve this problem, let $E_n = \{\text{no 5 or 7 appears on first } n-1 \text{ rolls, and 5 appears on } n\text{-th roll}\}$.
- Then $E_1, E_2, ...$ are mutually exclusive.

Example (cont.)

- Now,

$$\begin{aligned} P[\text{roll a } 5] &= 4/36 = 1/9 \\ P[\text{roll a } 7] &= 6/36 \\ P[\text{not roll a 5 or } 7] &= 1 - 10/36 = 13/18 \end{aligned}$$

and

$$\begin{aligned} P[E_n] =& P[\{\text{no 5 or 7 on 1st roll}\} \cap \cdots \\ & \cdots \cap \{\text{no 5 or 7 on } n-1 \text{ roll}\} \cap \{\text{5 on } n\text{th roll}\}] \\ =& P[\{\text{no 5 or 7 on 1st roll}\}] \times \cdots \\ & \cdots \times P[\{\text{no 5 or 7 on } n-1 \text{ roll}\}] \times P[\{\text{5 on } n\text{th roll}\}] \end{aligned}$$

- Finally, we want

$$P[\cup_{n=1}^{\infty} E_n] = \sum_{n=1}^{\infty} P[E_n] =$$

$$= \frac{1}{9} \times \sum_{n=1}^{\infty} \left(\frac{13}{18}\right)^{n-1} = \frac{1}{9} \times \frac{1}{1 - 13/18}$$