Unit 2 Conditioning and Independence

Conditioning
- revising a model based on new information

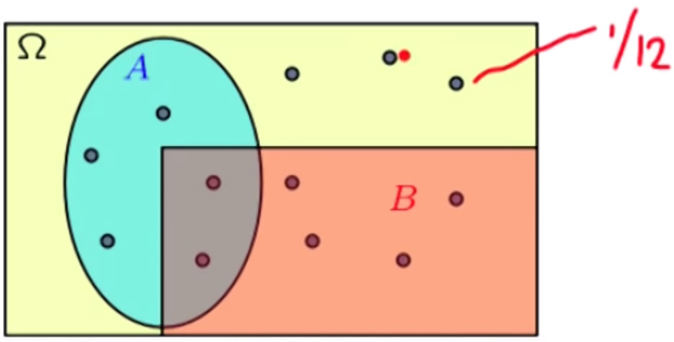
Independence
- assume outcomes are not related to each other to simplify complex models

<u>Conditioning:</u>

Conditional probability, 3 important rules
- multiplication rule
- total probability theorem
- Bayes' rule - foundation of inference theory

Use new information to revise a model

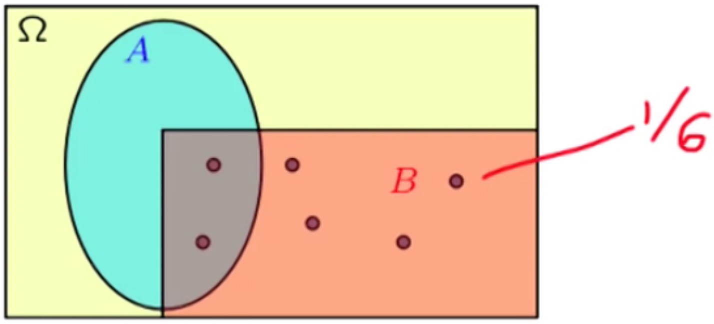
Assume 12 equally likely outcomes



$$\mathbf{P}(A) = \frac{5}{12} \qquad \mathbf{P}(B) = \frac{6}{12}$$

Probability of A given B has happened
P(A|B)

If told B occurred:



$\mathbf{P}(\underset{\sim}{A} \mid \underline{B}) = \frac{2}{6} = \frac{1}{3}$ $\quad \mathbf{P}(B \mid B) = 1$

Definition of conditional probability

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

defined only when $\mathbf{P}(B) > 0$

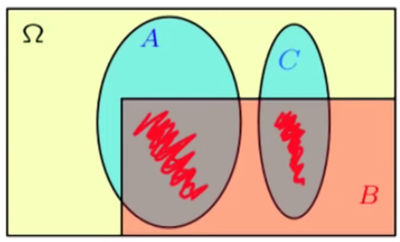
Conditional probabilities share properties of ordinary probabilities i.e. the same axioms apply
(for disjoint A and C below)

Conditional probabilities share properties of ordinary probabilities

$\mathbf{P}(A \mid B) \geq 0$ $\qquad$ assuming $\mathbf{P}(B) > 0$

$\mathbf{P}(\Omega \mid B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

$\mathbf{P}(B \mid B) = \frac{P(B \cap B)}{P(B)} = 1$



If $A \cap C = \emptyset$, then $\mathbf{P}(A \cup C \mid B) = \mathbf{P}(A \mid B) + \mathbf{P}(C \mid B)$

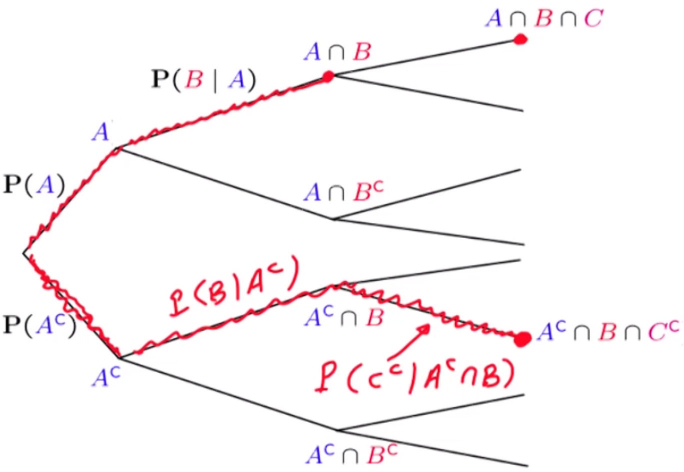
$= \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P(\overbrace{(A \cap B)} \cup \overbrace{(C \cap B)})}{P(B)} = \frac{P(A \cap B) + P(C \cap B)}{P(B)} =$

$= P(A \mid B) + P(C \mid B)$ $\quad$ also finite additivity
countable

The multiplication rule

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

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



$$P(\underline{A^c \cap B} \cap \underline{C^c}) =$$
$$= P(A^c \cap B)\, P(C^c \mid A^c \cap B)$$
$$= P(A^c) \cdot P(B \mid A^c)\, P(C^c \mid A^c \cap B)$$

$$P(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= P(A_1) \prod_{i=2}^{n} P(A_i \mid A_1 \cap \cdots \cap A_{i-1})$$

General form for multiplication rule is a product of the probability of A[1] with the probability that all previous events A[i] up to i=n have already occurred

1. $\mathbf{P}(A \cap B \cap C^c) = \mathbf{P}(A \cap B)\ \mathbf{P}(C^c \mid A \cap B)$

   True ✔ **Answer:** True

2. $\mathbf{P}(A \cap B \cap C^c) = \mathbf{P}(A)\ \mathbf{P}(C^c \mid A)\ \mathbf{P}(B \mid A \cap C^c)$

   True ✔ **Answer:** True

3. $\mathbf{P}(A \cap B \cap C^c) = \mathbf{P}(A)\ \mathbf{P}(C^c \cap A \mid A)\ \mathbf{P}(B \mid A \cap C^c)$

   False ✖ **Answer:** True

4. $\mathbf{P}(A \cap B \mid C) = \mathbf{P}(A \mid C)\ \mathbf{P}(B \mid A \cap C)$

   False ✖ **Answer:** True

Solution:

1. True. This is the usual multipication rule applied to the two events $A \cap B$ and C^c.
2. True. This is the usual multiplication rule.
3. True. This is because

$$\mathbf{P}(C^c \cap A \mid A) = \frac{\mathbf{P}(C^c \cap A \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(C^c \cap A)}{\mathbf{P}(A)} = \mathbf{P}(C^c \mid A).$$

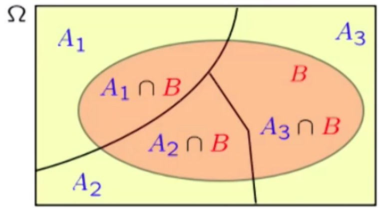
   So, this statement is equivalent to the one in part 2.

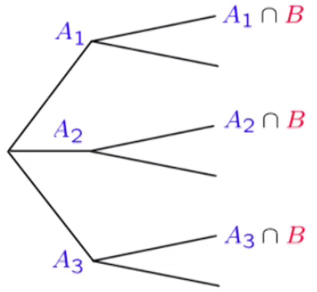
4. True. This is the usual multiplication rule $\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B \mid A)$, applied to a model/universe in which event C is known to have occurred.

Total Probability Theorem

- disjoint events that cover all possible outcomes (partition)
- if there were infinite partitions then it would be an infinite sum across all scenarios

Total probability theorem





- Partition of sample space into $A_1, A_2, A_3, \ldots$
- Have $\mathbf{P}(A_i)$, for every i
- Have $\mathbf{P}(B \mid A_i)$, for every i

$$\mathbf{P}(B) = \mathbf{P}(B \cap A_1) + \mathbf{P}(B \cap A_2) + \mathbf{P}(B \cap A_3)$$

$$= \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \cdots + \cdots$$

$$\sum_i \mathbf{P}(A_i) = 1$$

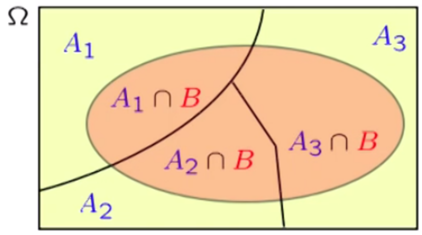
weights

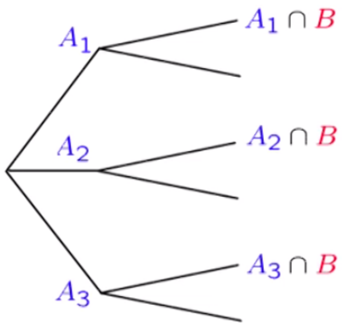
$$\boxed{\mathbf{P}(B) = \sum_i \mathbf{P}(A_i)\,\mathbf{P}(B \mid A_i)}$$

weighted average of $\mathbf{P}(B \mid A_i)$

Bayes' Rule

Bayes' rule





- Partition of sample space into A_1, A_2, A_3
- Have $\mathbf{P}(A_i)$, for every i initial "beliefs"
- Have $\mathbf{P}(B \mid A_i)$, for every i

revised "beliefs," given that B occurred:

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

$$\boxed{\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\sum_j \mathbf{P}(A_j)\mathbf{P}(B \mid A_j)}}$$

- Bayesian inference
  - initial beliefs $\mathbf{P}(A_i)$ on possible causes of an observed event B
  - model of the world under each A_i: $\mathbf{P}(B \mid A_i)$

$$A_i \xrightarrow[\mathbf{P}(B \mid A_i)]{\text{model}} B$$

  - draw conclusions about causes

$$B \xrightarrow[\mathbf{P}(A_i \mid B)]{\text{inference}} A_i$$

Independence:

In this coin tossing example the unconditional probability of getting heads in the second toss is the same as the conditional probability of getting it in the first (red)
This basically means that the events are independent, the result of the first toss has not impacted our beliefs in the result of the second toss

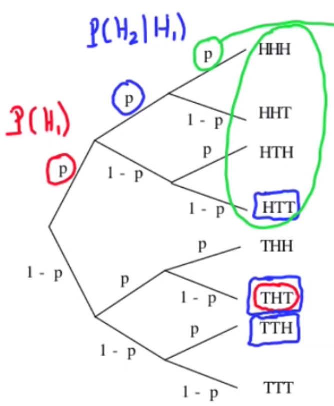
A model based on conditional probabilities

- 3 tosses of a biased coin: $\mathbf{P}(H) = p$, $\mathbf{P}(T) = 1 - p$

$P(H_2 \mid H_1) = p = P(H_2 \mid T_1)$

$P(H_2) = P(H_1)\,P(H_2 \mid H_1) + P(T_1)\,P(H_2 \mid T_1) = p$

$P(H_3 \mid H_1 \cap H_2)$



- Multiplication rule: $\mathbf{P}(THT) = (1-p)\,p\,(1-p)$
- Total probability:
  $\mathbf{P}(1 \text{ head}) = 3p(1-p)^2$
- Bayes rule:
  $\mathbf{P}(\text{first toss is H} \mid 1 \text{ head}) = \dfrac{P(H_1 \cap 1 \text{ head})}{P(1 \text{ head})}$

$$= \frac{p(1-p)^2}{3p(1-p)^2} = \frac{1}{3}$$

Independence of two events:
Intuitive definition P(B|A) = P(B) i.e. A provides no new information about B

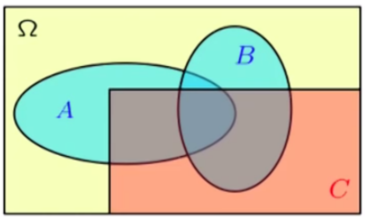
Definition of independence: $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$

Being independent is not at all like being disjoint
In fact disjoint events are not independent because knowing that A happens means B definitely did not happen, which gives us information about B

If A and B are independent then so are their complements

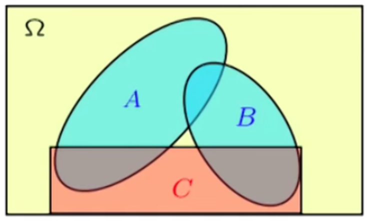
Conditional independence

- Conditional independence, given C, is defined as independence under the probability law $\mathbf{P}(\cdot \mid C)$



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

Assume A and B are independent



- If we are told that C occurred, are A and B independent? No

Because once C has occurred then only one or none of A or B can occur so we have some new information on them

Independence vs conditional independence

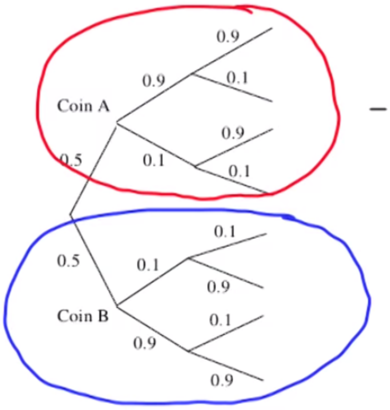
In the below example of 2 coins being tossed with different probabilities of heads we calculate the unconditional probability of heads as 0.5. Meaning the average of getting heads is 0.5 (between the coins)
But if we get 10 heads in a row the likelihood of that occurrence with coin B is so small that it tells us something about which coin it is more likely to be
There is a difference in the conditional and unconditional probability therefore there is not independence between the coin tosses

Conditioning may affect independence

given a coin: independent tosses

- Two unfair coins, A and B:
  $\mathbf{P}(H \mid \text{coin } A) = 0.9$, $\mathbf{P}(H \mid \text{coin } B) = 0.1$

- choose either coin with equal probability



- Are coin tosses independent?

No!

– Compare:
$\mathbf{P}(\text{toss } 11 = H) = \mathbb{P}(A)\,\mathbb{P}(H_{11} \mid A) + \mathbb{P}(B)\,\mathbb{P}(H_{11} \mid B)$
$= 0.5 \times 0.9 + 0.5 \times 0.1 = 0.5$

$\mathbf{P}(\text{toss } 11 = H \mid \text{first 10 tosses are heads})$

$\approx \mathbb{P}(H_{11} \mid A) = 0.9$

Independence of a collection of events

Information on some events does not change probabilities related to the remaining events

$$A_1, A_2, \ldots, \text{indep} \Rightarrow \mathbb{P}(A_3 \cap A_4^c) = \mathbb{P}(A_3 \cap A_4^c \mid A_1 \cup (A_2 \cap A_5^c))$$

Important to note that the indices are different on each side of the equality

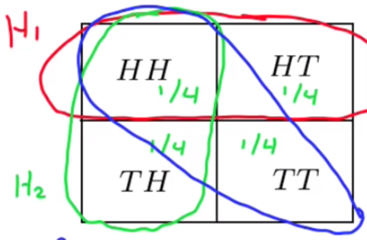
Definition: Events $A_1, A_2, \ldots, A_n$ are called **independent** if:
$\mathbf{P}(A_i \cap A_j \cap \cdots \cap A_m) = \mathbf{P}(A_i)\mathbf{P}(A_j)\cdots\mathbf{P}(A_m)$ for any distinct indices $i, j, \ldots, m$

If all pairs are independent it does not necessarily mean the entire collection of events is independent

$$\left.\begin{aligned} \mathbf{P}(A_1 \cap A_2) &= \mathbf{P}(A_1) \cdot \mathbf{P}(A_2) \\ \mathbf{P}(A_1 \cap A_3) &= \mathbf{P}(A_1) \cdot \mathbf{P}(A_3) \\ \mathbf{P}(A_2 \cap A_3) &= \mathbf{P}(A_2) \cdot \mathbf{P}(A_3) \end{aligned}\right\} \quad \text{pairwise independence}$$

$$\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2) \cdot \mathbf{P}(A_3)$$

Independence vs. pairwise independence



- Two independent fair coin tosses
  - H_1: First toss is H
  - H_2: Second toss is H

  $\mathbf{P}(H_1) = \mathbf{P}(H_2) = 1/2$

- C: the two tosses had the same result $= \{HH, TT\}$

$P(H_1 \cap C) = P(H_1 \cap H_2) = 1/4$ $\quad P(H_1)P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ $\quad$ H_1, C: indep. H_2, C: indep.

$P(H_1 \cap H_2 \cap C) = P(HH) = 1/4$
$P(H_1)P(H_2)P(C) = 1/8$ $\Big\rangle$ diff.

$P(C \mid H_1) = P(H_2 \mid H_1) = P(H_2) = 1/2 = P(C)$

$P(C \mid H_1 \cap H_2) = 1 \neq P(C) = 1/2$

H_1, H_2, and C are pairwise independent, but not independent

The events are all pairwise independent
Knowing that H1 happened doesn't affect our probability of C happening (likewise for H2)
But if we know that H1 and H2 happened then the probability of C happening is now 100% when previously it was 50%

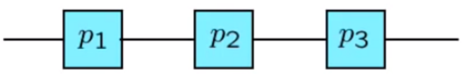
Reliability
Probability that a 'unit' is operational/ running (servers for example)

Reliability

p_i: probability that unit i is "up"

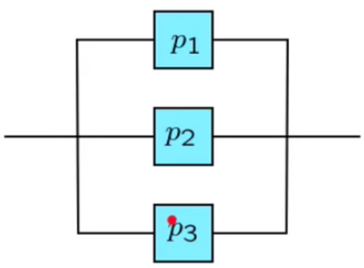
independent units

U_i: ith unit up
$U_1, U_2, \ldots, U_n$ independent
F_i: ith unit down
$\Rightarrow F_i$ independent



probability that system is "up"?

$P(\text{system up}) = P(U_1 \cap U_2 \cap U_3)$
$= P(U_1)P(U_2)P(U_3) = p_1 p_2 p_3$



$P(\text{system is up}) = P(U_1 \cup U_2 \cup U_3)$
$= 1 - P(F_1 \cap F_2 \cap F_3)$
$= 1 - P(F_1)P(F_2)P(F_3)$
$= 1 - (1-p_1)(1-p_2)(1-p_3)$

So the reliability of the second system is better

Problems

A chess tournament problem. This year's Belmont chess champion is to be selected by the following procedure. Bo and Ci, the leading challengers, first play a two-game match. If one of them wins both games, he gets to play a two-game **second round** with Al, the current champion. Al retains his championship unless a second round is required and the challenger beats Al in both games. If Al wins the initial game of the second round, no more games are played.

Furthermore, we know the following:
- The probability that Bo will beat Ci in any particular game is 0.6.
- The probability that Al will beat Bo in any particular game is 0.5.
- The probability that Al will beat Ci in any particular game is 0.7.

Assume no tie games are possible and all games are independent.

1. Determine the a priori probabilities that
(a) the second round will be required.
(b) Bo will win the first round.
(c) Al will retain his championship this year.

2. Given that the second round is required, determine the conditional probabilities that
(a) Bo is the surviving challenger.
(b) Al retains his championship.

3. Given that the second round was required and that it comprised only one game, what is the conditional probability that it was Bo who won the first round?