5005 CEM / 5014 CEM

29/1/2020
and 31/1/2020

1.2 Conditional Probability

last week

Recap —

- $0 \leq P(A) \leq 1$ for each event $A \subseteq S$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any events A, B
- If A and B are mutually exclusive then $A \cap B = \emptyset$
- $P(\overline{A}) = 1 - P(A)$ for any event A

Example 1.8

- If a single fair die is tossed, then

$$P(\text{face 2 turns up}) = \frac{1}{6}$$

- If a single fair die is tossed, and it is known that the face that turns up is an even number, then

$$P(\text{face 2 turns up}) = \frac{1}{3}$$

- If a single fair die is tossed, and it is known that the face that turns up is an odd number, then

$$P(\text{face 2 turns up}) = \frac{0}{3} = 0$$

$\square$

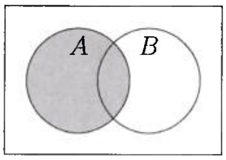
Notes —

- The probability of an event is **conditioned** by what other events we know to have occurred.
- The **sample space** that a probability is calculated with respect to has a great bearing upon the probability.

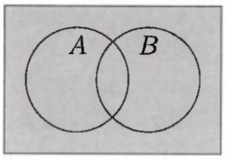
The **probability** of event A is

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } S}$$

assume equally likely outcomes



divided by



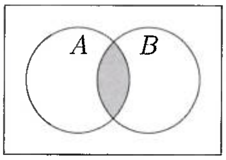
The **probability** of event A **conditional** on knowing that event B has occurred is

IDEA:

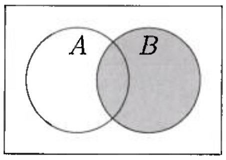
$$P(A\,|\,B) = \frac{\text{number of outcomes in } A \cap B}{\text{number of outcomes in } B}$$

→ event is $A \cap B$

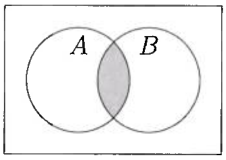
→ sample space is B



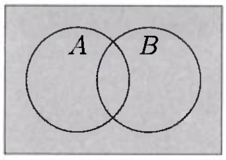
divided by

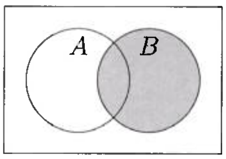


which is equivalent to

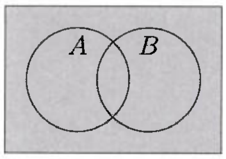


divided by





divided by



In general:

definition

■ The **conditional probability of A given B** is

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

event of interest

has already occurred (additional information)

Example 1.9

An electronic display is equally likely to show any of the digits 1,2,3,4,5,6,7,8,9.

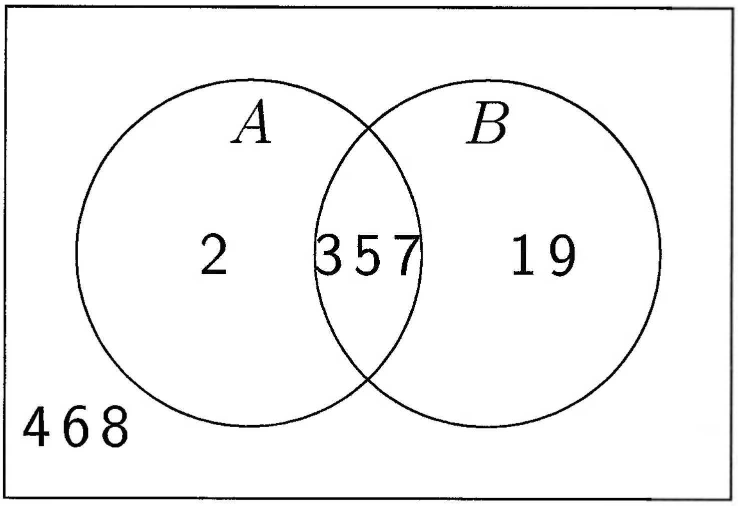
Determine the probability that it shows a prime number (i.e. one of 2, 3, 5 and 7):

(i) given no knowledge about the number

(ii) given the information that the number is odd

Solution.
Let A be the event "a prime number".

Let B be the event "an odd number".



(i) $P(A) = \frac{|A|}{|S|} = \frac{4}{9}$ $= P(\text{prime})$

(ii) $P(A \mid B) = \frac{|A \cap B|}{|B|} = \frac{3}{5}$ $= P(\text{prime} \mid \text{odd})$

<u>Note</u> : $P(\text{prime} \cap \text{odd}) = \frac{3}{9}$ $\square$

AND

Addition rule (always):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

1.2.1 Independent Events

Two events are **independent** if the occurrence of either event does not change the probability of the other event occurring.

definition

■ Events A and B are **independent** if

$$P(A \,|\, B) \;=\; P(A)$$

(or equivalently)

$$P(B \,|\, A) \;=\; P(B)$$

Notes —

- *Physically* independent events are always *statistically* independent.
- Since $\boxed{P(A \,|\, B) = \dfrac{P(A \cap B)}{P(B)}}$ we have → definition of conditional probability

$$\boxed{P(A \cap B) \;=\; P(A \,|\, B) \times P(B)}$$ → often useful in general case

- If A and B are independent then $P(A \,|\, B) = P(A)$ so ...

■ *Multiplication Law of Probability*

If A and B are **independent** events then

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

subsets of outcomes

Practice Problem. Two events A and B are such that

$$P(A) = 0.5 \qquad P(B) = 0.4 \qquad P(A \mid B) = 0.3$$

(a) Are A and B independent?

If independent then equal

$\boxed{\text{No}}$ since $P(A) \neq P(A \mid B)$

(b) Find the value of $P(A \cap B)$.

$$P(A \cap B) = P(A \mid B) \times P(B)$$
$$= 0.3 \times 0.4$$
$$= 0.12 \quad \rightarrow A \cap B \neq \emptyset$$

$\square$

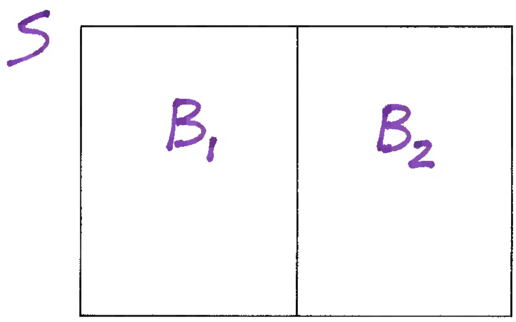
Note : $P(A) \times P(B) = 0.5 \times 0.4 = 0.2$

1.2.2 Law of Total Probability

no overlap

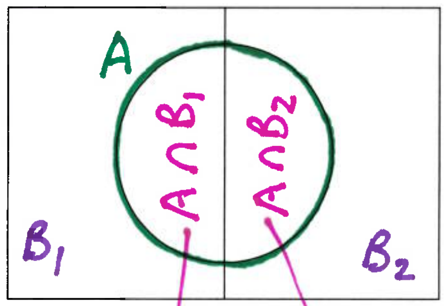
Suppose $S = B_1 \cup B_2$ where $B_1 \cap B_2 = \emptyset$ take up all of S

i.e., B_1 and B_2 are mutually exclusive and exhaustive (see page 1012).



partition of S into two parts

Now consider event A.



It is clear that

$$A = (A \cap B_1) \cup (A \cap B_2)$$

and that

$$(A \cap B_1) \cap (A \cap B_2) = \emptyset$$

i.e., A can be expressed as the union of two mutually exclusive events $(A \cap B_1)$ and $(A \cap B_2)$.

Therefore (using the addition law)

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) \\ &= P(A \,|\, B_1) \times P(B_1) + P(A \,|\, B_2) \times P(B_2) \end{aligned}$$

$P(\text{do well } M \mid \text{do well } P) = 0.8$

Example 1.10

- Of those students who do well in Physics, 80% also do well in Mathematics.
- Of those students who do not do well in Physics, only 30% do well in Mathematics. $P(\text{do well } M \mid \text{do } \underline{\text{not}} \text{ well } P) = 0.3$
- If 40% of students do well in Physics, what proportion do well in Mathematics? $P(\text{do well } P) = 0.4$

Solution.
Let A be the event "does well in Mathematics".
Let B_1 be the event "does well in Physics".
Let B_2 be the event "does not do well in Physics".

formalise

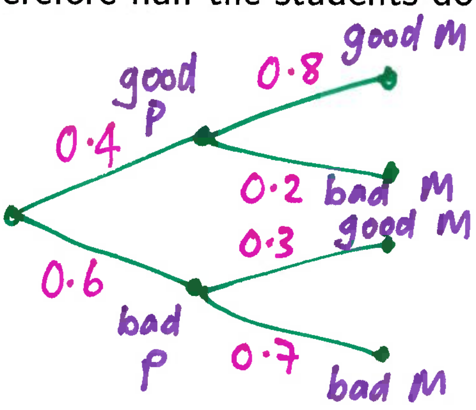
B_1 and B_2 are mutually exclusive and exhaustive events

So $P(M) = P(M|P) \times P(P) + P(M|\bar{P}) \times P(\bar{P})$

$$
\begin{aligned}
P(A) &= P(A \mid B_1) \times P(B_1) + P(A \mid B_2) \times P(B_2) \\
&= 0.8 \times 0.4 + 0.3 \times 0.6 \\
&= 0.32 + 0.18 \\
&= 0.5
\end{aligned}
$$

Therefore half the students do well in Mathematics.

Probability tree



$P(\text{good } P \cap \text{good } M) = 0.4 \times 0.8 = \boxed{0.32}$

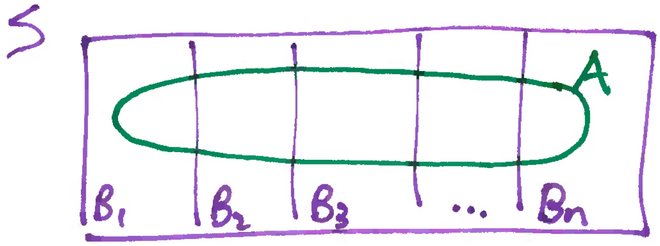
$P(\text{good } P \cap \text{bad } M) = 0.4 \times 0.2 = 0.08$

$P(\text{bad } P \cap \text{good } M) = 0.6 \times 0.3 = \boxed{0.18}$

$P(\text{bad } P \cap \text{bad } M) = 0.6 \times 0.7 = 0.42$

1.00

$P(\text{good } M) = 0.32 + 0.18 = \boxed{0.50}$



→ partition of S

■ *Law of Total Probability*

Suppose that $B_1, B_2, \ldots, B_n$ are mutually exclusive and exhaustive events, i.e.,

- $B_i \cap B_j = \emptyset$ for every i and j where $i \neq j$
- $B_1 \cup B_2 \cup \cdots \cup B_n = S$

Then

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n) \\ &= P(A \,|\, B_1) \times P(B_1) \\ &\quad + P(A \,|\, B_2) \times P(B_2) \\ &\quad + \ldots \\ &\quad + P(A \,|\, B_n) \times P(B_n) \end{aligned}$$

Horrific! → use probability tree instead

→ multiply probabilities along the branches

1.2.3 Tree Diagrams

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

Since $B \cap A = A \cap B$ this can be rearranged as

$$P(A \cap B) = P(A) \times P(B\,|\,A)$$ → "chain" rule

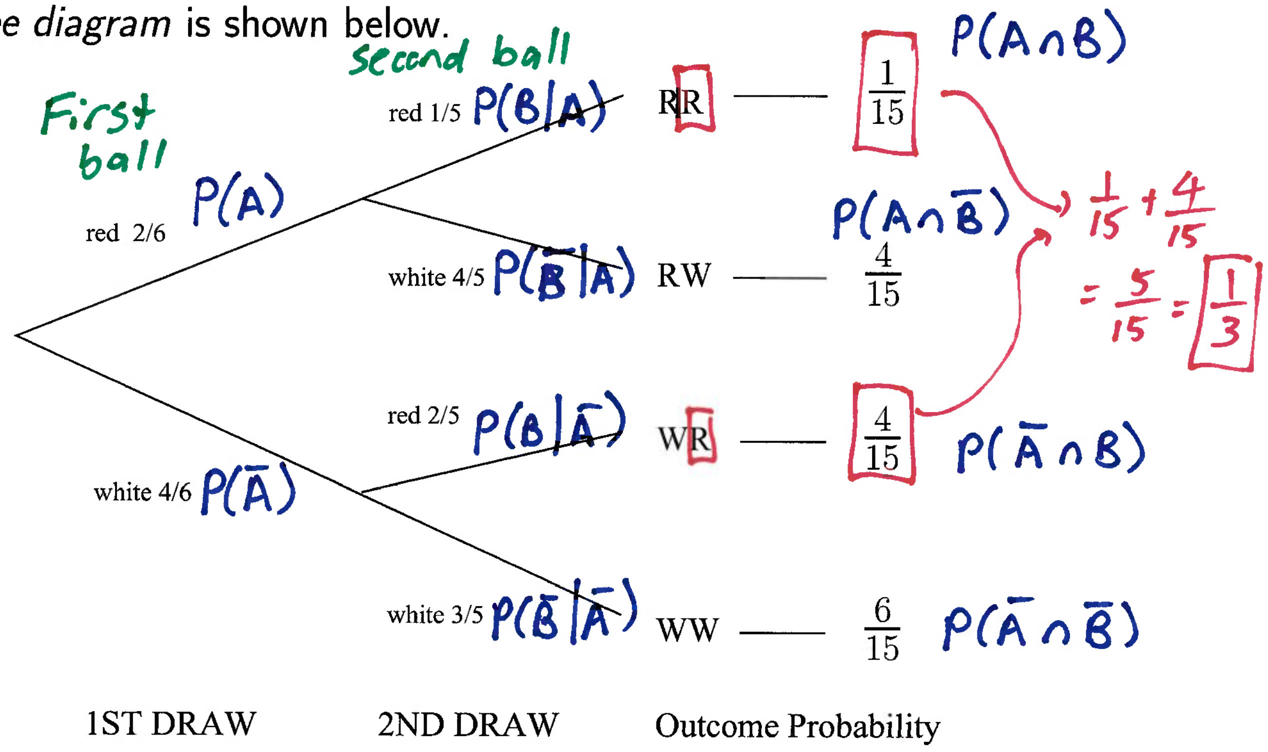
Repeating this idea gives the **chain rule** for conditional probabilities

$$P(A \cap B \cap C) = P(A) \times P(B\,|\,A) \times P(C\,|\,A \cap B)$$

Example 1.11 A bucket contains 2 red balls and 4 white balls. Two balls are drawn in sequence, *without replacement*. What is the probability that the second ball drawn is red?

A = First ball red.
B = Second ball red.

Solution.
A *tree diagram* is shown below.



pr003

- The probabilities along the branches are the conditional probabilities for the next stage, given the results of the previous stages.
- Multiplying the probabilities along each branch gives

(by the **chain rule**)

the probability of the outcome that branch represents.

$$\begin{aligned} P(RR) &= \tfrac{2}{6} \times \tfrac{1}{5} = \tfrac{1}{15} \\ P(RW) &= \tfrac{2}{6} \times \tfrac{4}{5} = \tfrac{4}{15} \\ P(WR) &= \tfrac{4}{6} \times \tfrac{2}{5} = \tfrac{4}{15} \\ P(RR) &= \tfrac{4}{6} \times \tfrac{3}{5} = \tfrac{6}{15} \end{aligned}$$

- Check that the total probability is 1, i.e.,

$$\tfrac{1}{15} + \tfrac{4}{15} + \tfrac{4}{15} + \tfrac{6}{15} = \tfrac{1+4+4+6}{15} = \tfrac{15}{15} = 1$$

- Then add the probabilities contributing to the event of interest, i.e.,

$$\begin{aligned} P(\text{second ball is red}) &= P(RR) + P(WR) \\ &= \tfrac{2}{6} \times \tfrac{1}{5} + \tfrac{4}{6} \times \tfrac{2}{5} \\ &= \tfrac{1}{15} + \tfrac{4}{15} = \tfrac{5}{15} = \boxed{\tfrac{1}{3}} \end{aligned}$$

$\square$

1.2.4 Bayes' Theorem

Thomas Bayes (1701–1761)
18th century British clergyman and mathematician
http://www.bayesian.org/bayesian/bayes.html

Idea of conditional probability —

Given that event B has happened in the past, what is the probability that event A will occur? $P(A \mid B)$

$P(\text{wet} \mid \text{forecast rain})$

Reverse question —

Given that event A has just occurred, what is the probability that it was preceded by the event B? $P(B \mid A)$

$P(\text{forecast rain} \mid \text{wet})$

definition of conditional probability

- Since $\boxed{P(A\,|\,B) = \dfrac{P(A \cap B)}{P(B)}}$ we have

$$P(A \cap B) \;=\; P(A\,|\,B) \times P(B)$$

same

- Since $\boxed{P(B\,|\,A) = \dfrac{P(B \cap A)}{P(A)}}$ we have

$$P(B \cap A) \;=\; P(B\,|\,A) \times P(A)$$

- But $P(A \cap B) = P(B \cap A)$ so we have

$$\boxed{P(A\,|\,B) \times P(B) \;=\; P(B\,|\,A) \times P(A)}$$

■ *Bayes' Theorem*

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

Sometimes called "Bayesian inversion"

Terminology (in machine learning) —

- $P(B)$ is known as the **prior** or **hypothesis** (before evidence is seen)
- $P(A)$ is known as the **evidence**
- $P(A\,|\,B)$ is known as the **likelihood** (indicates the compatibility of the evidence with the hypothesis)
- $P(B\,|\,A)$ is known as the **posterior** (after evidence is seen)

used in Forensic Science
(and legal cases)

H or T   H or H

Example 1.12

An engineer has a fair coin and a double-headed coin. She chooses one of the coins at random and tosses it. She obtains a head. Determine the probability that the coin that she tossed was double-headed.

Solution.

Let A be the event that "a head is obtained".

Let B_1 be the event that "the fair coin was chosen".

Let B_2 be the event that "the double-headed coin was chosen".

B_1 and B_2 are mutually exclusive and exhaustive events.

We know

$$P(B_1) = \tfrac{1}{2} \qquad P(B_2) = \tfrac{1}{2} \qquad P(A \mid B_1) = \tfrac{1}{2} \qquad P(A \mid B_2) = 1$$

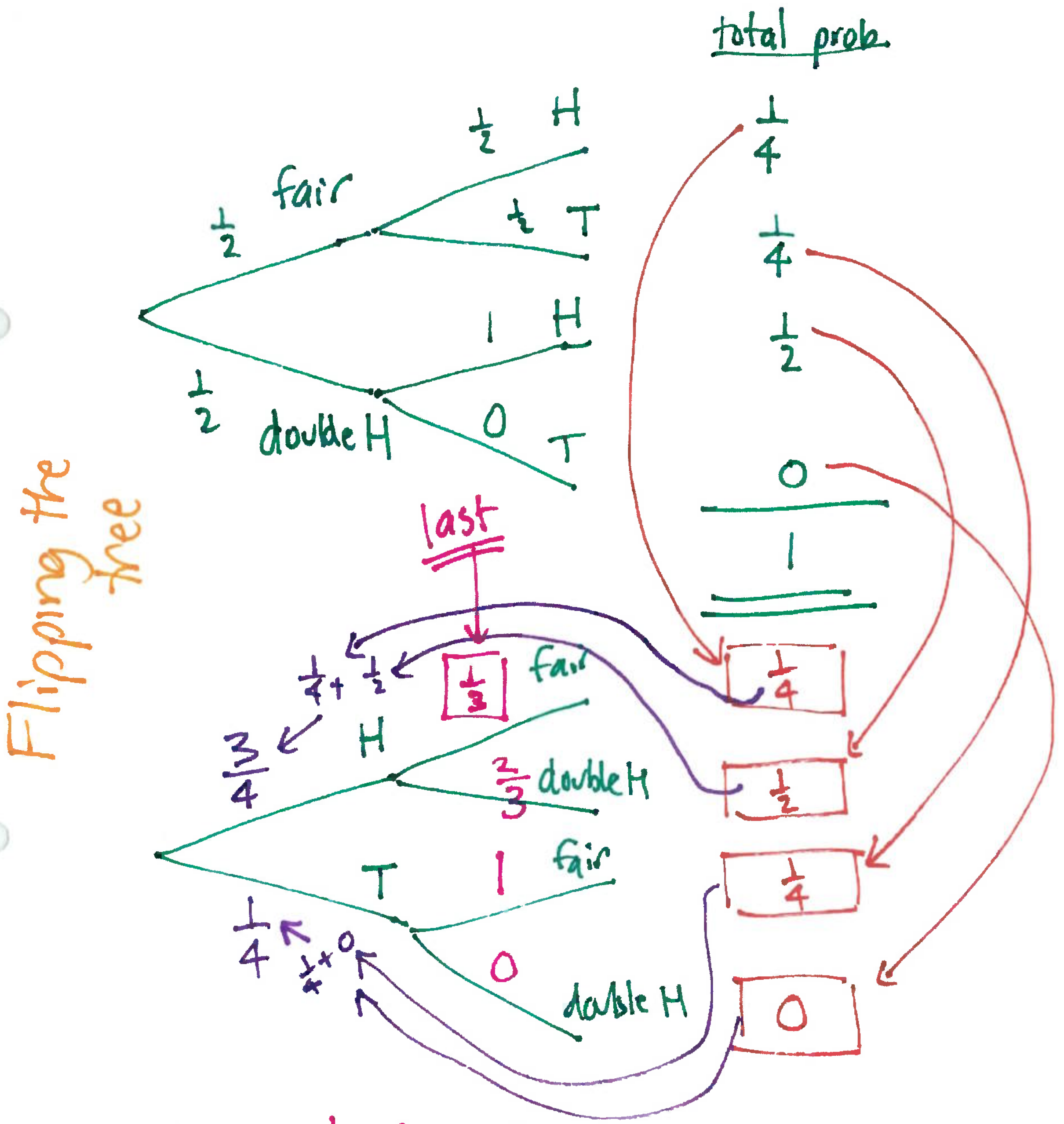
$$\begin{aligned} P(A) &= P(A \mid B_1) \times P(B_1) + P(A \mid B_2) \times P(B_2) \\ &= \tfrac{1}{2} \times \tfrac{1}{2} + 1 \times \tfrac{1}{2} \;=\; \tfrac{3}{4} \end{aligned}$$

$$\begin{aligned} P(\text{coin was double-headed}) &= P(B_2 \mid A) \\ &= \frac{P(A \mid B_2) \times P(B_2)}{P(A)} \\ &= \frac{1 \times \frac{1}{2}}{\frac{3}{4}} \;=\; \frac{2}{3} \end{aligned}$$

Bayes' theorem

► *A good thing about Bayes' Theorem is that, once the events have been carefully defined, we do not need to think!*

► We can often perform all the calculations on two probability trees.



$P(\text{fair} \mid H) = \frac{1}{3}$

$P(\text{double H} \mid H) = \frac{2}{3}$ ← solution to the problem

$P(\text{fair} \mid T) = 1$

$P(\text{double H} \mid T) = 0$

Summary

- The *conditional probability* of event A, given that event B is known to have occurred, is

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

then we can also write

$$P(A \cap B) \;=\; P(A \mid B) \times P(B)$$

- Events A and B are *independent* if $P(A \mid B) = P(A)$

(or equivalently) $P(B \mid A) = P(B)$

(or equivalently) $P(A \cap B) = P(A) \times P(B)$

- *Law of total probability* —

If $B_1, B_2, \ldots, B_n$ are mutually exclusive and exhaustive then

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n) \\ &= P(A \mid B_1)\, P(B_1) + P(A \mid B_2)\, P(B_2) \\ &\qquad + \ldots + P(A \mid B_n)\, P(B_n) \end{aligned}$$

- *Chain rule* — multiply probability along the branches in tree diagrams

$$\begin{aligned} P(A \cap B) &= P(A) \times P(B \mid A) \\ P(A \cap B \cap C) &= P(A) \times P(B \mid A) \times P(C \mid A \cap B) \end{aligned}$$

- *Bayes' Theorem*

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

- Use probability tree wherever possible.
- "Flipping" the tree.