MTL108
Conditional probability and independence

Rahul Singh

In real-world scenarios, events often depend on specific conditions or prior knowledge. For example, the probability of rain might change if we know it's already cloudy, or the likelihood of a medical diagnosis might shift based on test results. Conditional probability provides a mathematical framework to quantify these dependencies, enabling more accurate predictions and decision-making in fields like statistics, machine learning, and risk analysis.

Definition 1 (Conditional Probability)**.** Let A and B be two events in a probability space with $\mathbb{P}(B) > 0$. The conditional probability of A given B, denoted $\mathbb{P}(A \mid B)$, is defined as

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

This represents the probability that event A occurs, given that event B has already occurred.

Example 1. Suppose we have a fair six-sided die. Let A be the event that the roll is even (i.e., 2, 4, or 6), so $\mathbb{P}(A) = 1/2$. Let B be the event that the roll is greater than 3 (i.e., 4, 5, or 6), so $\mathbb{P}(B) = 1/2$. The intersection $A \cap B$ is the event of rolling 4 or 6, so $\mathbb{P}(A \cap B) = 1/3$.
Thus,

$$\mathbb{P}(A \mid B) = \frac{1/3}{1/2} = \frac{2}{3}.$$

This means, given that the roll is greater than 3, the probability it is even is 2/3.

Example 2. Consider a family with two children, where each child is equally likely to be a boy (B) or a girl (G) with probability 1/2, and the gender of one child does not affect the other. It is known that at least one of them is a girl. What is the probability that both are girls, given this information? What if it is known that the elder child is a girl?

Solutions: The sample space for the genders of two children (elder, younger) is:

$$\Omega = \{(G,G),(G,B),(B,G),(B,B)\},$$

where G denotes a girl and B denotes a boy. Each outcome has probability:

$$\mathbb{P}((G,G)) = \mathbb{P}((G,B)) = \mathbb{P}((B,G)) = \mathbb{P}((B,B)) = \frac{1}{4}.$$

At Least One Girl: Let A be the event that both children are girls, i.e., $A = \{(G,G)\}$. Thus,

$$\mathbb{P}(A) = \frac{1}{4}.$$

Let B be the event that at least one child is a girl, i.e., $B = \{(G,G), (G,B), (B,G)\}$. The probability of B is:

$$\mathbb{P}(B) = \mathbb{P}((G,G)) + \mathbb{P}((G,B)) + \mathbb{P}((B,G)) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

The intersection $A \cap B$ is the event that both are girls and at least one is a girl, which is simply $A = \{(G,G)\}$, since (G,G) satisfies both conditions. Thus,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) = \frac{1}{4}.$$

Using the definition of conditional probability,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Alternatively, given at least one girl, the sample space reduces to $\{(G,G), (G,B), (B,G)\}$, each equally likely with conditional probability $1/3$. Only (G,G) corresponds to both being girls, so the probability is $1/3$.

Elder Child is a Girl: Let C be the event that the elder child is a girl, i.e., $C = \{(G,G), (G,B)\}$. The probability of C is:

$$\mathbb{P}(C) = \mathbb{P}((G,G)) + \mathbb{P}((G,B)) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

The intersection $A \cap C$ is the event that both are girls and the elder is a girl, which is $A = \{(G,G)\}$. Thus,

$$\mathbb{P}(A \cap C) = \mathbb{P}(A) = \frac{1}{4}.$$

Using conditional probability,

$$\mathbb{P}(A \mid C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Alternatively, given the elder child is a girl, the sample space reduces to $\{(G,G), (G,B)\}$, each with conditional probability $1/2$. Only (G,G) corresponds to both being girls, so the probability is $1/2$. The difference in probabilities ($1/3$ vs. $1/2$) arises because the condition "at least one is a girl" includes more outcomes ((B,G)) than "elder is a girl," affecting the conditional sample space. This highlights how specific information influences conditional probabilities, a key insight in applications related to real-world scenarios.

Definition 2 (Two independence events)**.** Two events A and B are said to be independent if

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

Equivalently, if $\mathbb{P}(B) > 0$, then $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, meaning the occurrence of B does not affect the probability of A.

Definition 3 (Mutually independent events)**.** For more than two events, say $A_1, A_2, \ldots, A_n$, they are mutually independent if for every subset $I \subseteq \{1, 2, \ldots, n\}$,

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Definition 4 (Pairwise independent events)**.** A collection of events, say $A_1, A_2, \ldots, A_n$, are pairwise independent if for every $1 \leq i \neq j \leq n$,

$$P(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

Example 3. Consider flipping two fair coins. Let A be the event that the first coin is heads, $\mathbb{P}(A) = 1/2$. Let B be the event that the second coin is heads, $\mathbb{P}(B) = 1/2$. The intersection $A \cap B$ (both heads) has $\mathbb{P}(A \cap B) = 1/4 = (1/2)(1/2)$, so A and B are independent.
Next, let C be the event that both coins show the same face and D be the event denoting at least one head. Then $\mathbb{P}(C) = 1/2$, $\mathbb{P}(D) = 3/4$, and $\mathbb{P}(C \cap D) = \mathbb{P}(\text{both heads}) = 1/4 \neq (1/2)(3/4) = \mathbb{P}(C)\mathbb{P}(D)$, so C and D are not independent.

Remark 1 (Independent vs disjoint events)**.** In probability theory, the concepts of *independent* events and *disjoint* events are distinct and often confused, but they describe fundamentally different relationships between events. Understanding their differences is crucial for correctly applying probability concepts. In summary, independence relates to the probabilistic influence of events, while disjointness concerns their mutual exclusivity.

Recall, two events A and B are disjoint (or mutually exclusive) if they cannot occur simultaneously, i.e., $A \cap B = \emptyset$. Thus, $\mathbb{P}(A \cap B) = 0$.

Disjoint events with positive probabilities cannot be independent, except in the trivial case where at least one event has zero probability. If A and B are disjoint and $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$, then $\mathbb{P}(A \cap B) = 0$, but $\mathbb{P}(A)\mathbb{P}(B) > 0$, so $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$, violating the condition for independence. Conversely, independent events with positive probabilities are typically not disjoint, as their intersection has probability $\mathbb{P}(A)\mathbb{P}(B) > 0$. For instance, in the coin flip example, the events are independent and their intersection (both heads) has positive probability.

Lemma 1. *For any two events A and B,*

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

Proof. If $\mathbb{P}(A) = 0$, then

$$A \cap B \subseteq A \Rightarrow 0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(A) \Rightarrow \mathbb{P}(A \cap B) = 0 = \mathbb{P}(B \mid A)\mathbb{P}(A).$$

Next, if $\mathbb{P}(A) > 0$, then by definition of conditional probability:

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

□

Theorem 1 (Total probability theorem)**.** *If $\{A_i\}_{i=1}^n$ is a partition of the sample space (mutually exclusive and exhaustive events) and $\mathbb{P}(B) > 0$, then*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).$$

Proof. Using Theorem 3 in Topic-2 notes, we have

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i).$$

Next using Lemma 1, we have $\mathbb{P}(B \cap A_i) = (B \mid A_i)\mathbb{P}(A_i)$. Consequently,

$$\sum_{i=1}^{n} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).$$

□

Theorem 2 (Bayes' rule or Bayes' theorem)**.** *For events A and B with $\mathbb{P}(B) > 0$,*

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

More generally, if $\{A_i\}_{i=1}^{n}$ is a partition of the sample space (mutually exclusive and exhaustive events), then for any event B with $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A_k \mid B) = \frac{\mathbb{P}(B \mid A_k)\mathbb{P}(A_k)}{\sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.$$

Remark 2. Here $\mathbb{P}(A_k \mid B)$ is known as posterior probability; "posterior" means after updating based on the evidence.

Proof. By definition of conditional probability,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \tag{1}$$

provided $\mathbb{P}(B) > 0$. From Lemma 1, we have $\mathbb{P}(A \cap B) = \mathbb{P}(B \cap A) = \mathbb{P}(B \mid A)\mathbb{P}(A)$

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

This completes the proof of the first statement.
Second part: Collection $\{A_i\}_{i=1}^{n}$ is a partition of the sample space, that is, $\bigcup_{i=1}^{n} A_i = \Omega$ and $A_i \cap A_j = \emptyset$ for $i \neq j$. If $\mathbb{P}(B) > 0$, the law of total probability gives:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).$$

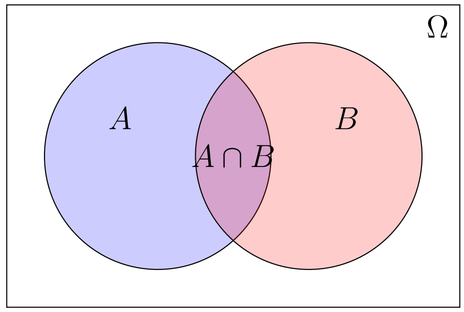
Thus, for any A_k,

$$\mathbb{P}(A_k \mid B) = \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A_k)\mathbb{P}(A_k)}{\sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.$$

This establishes the general form of Bayes' theorem. □

Venn Diagram Illustration

The diagram helps visualize how $\mathbb{P}(A \cap B)$ relates to $\mathbb{P}(A)$ and $\mathbb{P}(B)$, which is central to understanding Bayes' theorem.



In the diagram, the intersection $A \cap B$ is the region where the circles overlap, representing the joint event. Bayes' theorem relates the conditional probabilities by using this intersection and the marginal probabilities $\mathbb{P}(A)$ and $\mathbb{P}(B)$.

Example 4. Consider a disease affecting 1% of a population ($\mathbb{P}(D) = 0.01$). A test is 99% accurate: $\mathbb{P}(T^+ \mid D) = 0.99$ (true positive) and $\mathbb{P}(T^- \mid D^c) = 0.99$ (true negative), so $\mathbb{P}(T^+ \mid D^c) = 0.01$ (false positive), with $\mathbb{P}(D^c) = 0.99$.
To find $\mathbb{P}(D \mid T^+)$, first we compute the probability of having the disease given a positive test:

$$\begin{aligned}\mathbb{P}(T^+) &= \mathbb{P}(T^+ \mid D)\mathbb{P}(D) + \mathbb{P}(T^+ \mid D^c)\mathbb{P}(D^c) = (0.99)(0.01) + (0.01)(0.99) \\ &= 0.0099 + 0.0099 = 0.0198.\end{aligned}$$

Now,

$$\mathbb{P}(D \mid T^+) = \frac{\mathbb{P}(T^+ \mid D)\mathbb{P}(D)}{\mathbb{P}(T^+)} = \frac{(0.99)(0.01)}{0.0198} = \frac{0.0099}{0.0198} = 0.5.$$

Thus, there is a 50% chance of having the disease given a positive test, highlighting the impact of the low prior probability $\mathbb{P}(D)$.

Example 5. Two fair dice are rolled independently. Let $A = \{\text{Sum of dice roll is 6}\}$, $B = \{\text{outcome on first die is 4}\}$, and $C = \{\text{Sum of dice roll is 7}\}$.

1. Check whether A and B are independent events.
2. Check whether C and B are independent events.

You will find that the events A and B are dependent (not independent), but the events C and B are independent. Can you explain why?

Solution: The outcomes where the sum is 6 are: $(1,5), (2,4), (3,3), (4,2), (5,1)$. There are 5 outcomes, so

$$\mathbb{P}(A) = \frac{5}{36}.$$

The outcomes where the first die is 4 are: $(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)$. There are 6 outcomes, so

$$\mathbb{P}(B) = \frac{6}{36} = \frac{1}{6}.$$

The intersection $A \cap B$ is the event where the first die is 4 and the sum is 6. The only outcome is $(4,2)$, since $4+2=6$. Thus,

$$\mathbb{P}(A \cap B) = \frac{1}{36}.$$

To check independence, compute:

$$\mathbb{P}(A)\mathbb{P}(B) = \frac{5}{36} \cdot \frac{1}{6} = \frac{5}{216}.$$

Since

$$\mathbb{P}(A \cap B) = \frac{1}{36} = \frac{6}{216} \neq \frac{5}{216} = \mathbb{P}(A)\mathbb{P}(B),$$

the events A and B are not independent, so they are dependent.
Alternatively, compute the conditional probability:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/36}{1/6} = \frac{1}{6}.$$

Compare with $\mathbb{P}(A)$:

$$\mathbb{P}(A) = \frac{5}{36} \neq \frac{1}{6} = \mathbb{P}(A \mid B).$$

Since $\mathbb{P}(A \mid B) \neq \mathbb{P}(A)$, the occurrence of B affects the probability of A, confirming dependency.

Next, the outcomes where the sum is 7 are: $(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$. There are 6 outcomes, so

$$\mathbb{P}(C) = \frac{6}{36} = \frac{1}{6}.$$

The intersection $C \cap B$ is the event where the first die is 4 and the sum is 7. The only outcome is $(4,3)$, since $4+3=7$. Thus,

$$\mathbb{P}(C \cap B) = \frac{1}{36}.$$

Compute:

$$\mathbb{P}(C)\mathbb{P}(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

Since

$$\mathbb{P}(C \cap B) = \frac{1}{36} = \mathbb{P}(C)\mathbb{P}(B),$$

the events C and B are independent.
Compute the conditional probability:

$$\mathbb{P}(C \mid B) = \frac{\mathbb{P}(C \cap B)}{\mathbb{P}(B)} = \frac{1/36}{1/6} = \frac{1}{6}.$$

Since

$$\mathbb{P}(C) = \frac{1}{6} = \mathbb{P}(C \mid B),$$

the occurrence of B does not affect the probability of C, confirming independence.
The events "sum of dice roll equals 6" and "first die equals 4" are dependent because knowing the first die is 4 changes the probability of the sum being 6 from 5/36 to 1/6. In contrast, the events "sum of dice roll equals 7" and "first die equals 4" are independent because the probability of the sum being 7 remains 1/6 regardless of whether the first die is 4.

Lemma 2. *Pairwise independence does not imply mutual independence.*

Proof. We prove this by using a counterexample. Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a probability space, where $\Omega = \{1,2,3,4\}$, $\mathcal{C} = 2^{\Omega}$ and

$$\mathbb{P}(\{x\}) = \begin{cases} 1/4 & \text{if } x = 1,2,3,4 \\ 0 & \text{otherwise.} \end{cases}$$

Consider events $E_1 = \{1,2\}$, $E_2 = \{2,3\}$, and $E_3 = \{1,3\}$. Then,

- $\mathbb{P}(E_1) = \mathbb{P}(E_2) = \mathbb{P}(E_3) = 1/2$
- $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1 \cap E_3) = \mathbb{P}(E_2 \cap E_3) = 1/4$
- $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2)$, $\mathbb{P}(E_1 \cap E_3) = \mathbb{P}(E_1)\mathbb{P}(E_3)$, and $\mathbb{P}(E_2 \cap E_3) = \mathbb{P}(E_2)\mathbb{P}(E_3)$
- $\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(\emptyset) = 0 \neq \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3)$.

Therefore, events E_1, E_2, and E_3 are pairwise independent but not mutually independent. □

Theorem 3. *If A and B are independent events then*

1. *A^c and B are independent events;*
2. *A and B^c are independent events;*
3. *A^c and B^c are independent events.*

Proof. We prove part 1 only; others similarly follow. Observe that,

$$\begin{aligned} \mathbb{P}(A^c \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B), \text{ because } A \text{ and } B \text{ are independent} \\ &= \mathbb{P}(B)(1 - \mathbb{P}(A)) \\ &= \mathbb{P}(B)\mathbb{P}(A^c), \text{ because } \mathbb{P}(A^c) = 1 - \mathbb{P}(A). \end{aligned}$$

Therefore, A^c and B are independent events. □

Theorem 4. *If A, B and C are independent events then A^c, B and C are independent events.*

Proof. Observe that,

$$\begin{aligned} \mathbb{P}(A^c \cap B \cap C) &= \mathbb{P}(A^c \cap (B \cap C)), \text{ using associative law of intersection} \\ &= \mathbb{P}(A^c)\mathbb{P}(B \cap C), \text{ using independence of } A \text{ and } B \cap C, \text{ and Theorem 1} \\ &= \mathbb{P}(A^c)\mathbb{P}(B)\mathbb{P}(C). \end{aligned}$$

Therefore, A^c, B and C are independent events. □

Conditional probability measure

Definition 5. Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a probability space. The conditional probability measure with respect to event $B \in \mathcal{C}$ where $\mathbb{P}(B) > 0$ is a function $\mathbb{P}_B(A) =: \mathcal{C} \to [0,1]$ defined by

$$\mathbb{P}_B(A) = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ for all } A \in \mathcal{C}.$$

$\mathbb{P}_B$ is known as a conditional probability measure on $(\Omega, \mathcal{C})$, and $(\Omega, \mathcal{C}, \mathbb{P}_B)$ is known as the corresponding conditional probability space.

Theorem 5. *A conditional probability measure is a probability measure.*

Proof. We assume the setup in the above definition. To complete the proof, we need to verify three axioms. Verifications of the first two axioms are straightforward.

Axiom 1 (Non-negativity) $\mathbb{P}_B(A) = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \geq 0$ for all $A \in \mathcal{C}$; non-negativity holds.

Axiom 2 (Normalization) $\mathbb{P}_B(\Omega) = \mathbb{P}(\Omega \mid B) = \frac{\mathbb{P}(\Omega\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$; normalization holds.

Axiom 3 (Countable additivity) Consider a collection of pairwise disjoint events $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{C}$ (i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$),

$$\begin{aligned}\mathbb{P}_B\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \frac{\mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty}(A_i \cap B)\right)}{\mathbb{P}(B)}, \text{ using distributive property of intersection over union.}\end{aligned}$$

Now, for $i \neq j$, $A_i \cap A_j = \emptyset$ implies that $(A_i \cap B) \cap (A_j \cap B) = \emptyset$ for $i \neq j$. So, $\{A_i \cap B\}_{i=1}^{\infty} \subseteq \mathcal{C}$ is a collection of pairwise disjoint events. Consequently, the third axiom for probability space $(\Omega, \mathcal{C}, \mathbb{P})$ gives

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty}(A_i \cap B)\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B).$$

Substituting the above in the first expression, we have

$$\mathbb{P}_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \mathbb{P}_B(A_i).$$

So,

$$\mathbb{P}_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}_B(A_i).$$

Consequently, the third axiom also holds.

Therefore, all three axioms for probability measure are satisfied by the conditional probability measure, so a conditional probability measure is a probability measure. □

Multiple conditioning:

Theorem 6. *For events A, B and C with $\mathbb{P}(A \cap B) > 0$,*

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

Proof. First notice that $\mathbb{P}(A) > \mathbb{P}(A \cap B) > 0$ and $\mathbb{P}(B) > \mathbb{P}(A \cap B) > 0$. Using the definition of conditional probability,

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

Using Lemma 1, we have

$$\mathbb{P}(C \cap A \cap B) = \mathbb{P}(C \cap A \mid B)\mathbb{P}(B) \text{ and } \mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B).$$

Substituting in the first expression, we obtain

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

□

Lemma 3. *For three events A, B and C with $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$ and $\mathbb{P}(C) > 0$, we have*

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \mid B \cap C)\, \mathbb{P}(B \mid C)\, \mathbb{P}(C).$$

Proof. Consider $D = B \cap C$ as an event, and applying Lemma 1, we have

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap D) = \mathbb{P}(A \mid D)\, \mathbb{P}(D).$$

Putting $D = B \cap C$ we get

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \mid B \cap C)\, \mathbb{P}(B \cap C).$$

Again applying Lemma 1 for $\mathbb{P}(B \cap C)$ and putting in the above expression we get

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \mid B \cap C)\, \mathbb{P}(B \mid C)\, \mathbb{P}(C).$$

□

Theorem 7 (Chain rule)**.** *Let $A_1, A_2, \ldots, A_n$ be the events such that $\mathbb{P}(A_i) > 0$ for $1 \leq i \leq n$. Then,*

$$\begin{aligned} &\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) \\ &= \mathbb{P}(A_1 \mid A_2 \cap \ldots \cap A_n)\, \mathbb{P}(A_2 \mid A_3 \cap \ldots \cap A_n)\, \mathbb{P}(A_3 \mid A_4 \cap \ldots \cap A_n) \ldots \mathbb{P}(A_{n-1} \mid A_n)\, \mathbb{P}(A_n). \end{aligned}$$

Monte Hall problem

A contestant is presented with three closed doors. Behind one of the doors is a car, and behind the other two doors are goats. The contestant initially selects one door at random.

After the initial choice, the host, who knows what is behind each door, opens one of the remaining two doors and reveals a goat. The host always opens a door with a goat and never opens the door chosen by the contestant.

The contestant is then given the option to either *stay* with the originally chosen door or *switch* to the other unopened door.

1. What is the probability of winning the car if the contestant stays with the original choice?
2. What is the probability of winning the car if the contestant switches to the other unopened door?

Solution:

Let's formalize this using probability. Assume the contestant always picks Door 1, and the car is equally likely to be behind any door, with probability $\frac{1}{3}$. The host opens a door revealing a goat (Door 2 or Door 3), and we calculate the probability that the car is behind Door 1 (stay) or Door 3 (switch, assuming the host opens Door 2).

Define events:

- C_i: The car is behind Door i ($i = 1, 2, 3$), with $P(C_i) = \frac{1}{3}$.
- H_2: The host opens Door 2, revealing a goat.

We want to compute:

- $P(C_1|H_2)$: Probability the car is behind Door 1 given the host opens Door 2 (stay).
- $P(C_3|H_2)$: Probability the car is behind Door 3 given the host opens Door 2 (switch).

Using Bayes' theorem, for $P(C_1|H_2)$:

$$P(C_1|H_2) = \frac{P(H_2|C_1)P(C_1)}{P(H_2)}.$$

Step 1: Compute probabilities.

- $P(C_1) = \frac{1}{3}$.
- $P(H_2|C_1)$: If the car is behind Door 1, the host chooses between Door 2 and Door 3 (both have goats) randomly, so $P(H_2|C_1) = \frac{1}{2}$.
- $P(H_2|C_2)$: If the car is behind Door 2, the host cannot open Door 2 (it has the car), so $P(H_2|C_2) = 0$.
- $P(H_2|C_3)$: If the car is behind Door 3, the host must open Door 2 (Door 1 is chosen, Door 3 has the car), so $P(H_2|C_3) = 1$.

Step 2: Compute $P(H_2)$.

$$P(H_2) = P(H_2|C_1)P(C_1) + P(H_2|C_2)P(C_2) + P(H_2|C_3)P(C_3).$$

$$P(H_2) = \left(\frac{1}{2} \cdot \frac{1}{3}\right) + \left(0 \cdot \frac{1}{3}\right) + \left(1 \cdot \frac{1}{3}\right) = \frac{1}{6} + 0 + \frac{1}{3} = \frac{1}{2}.$$

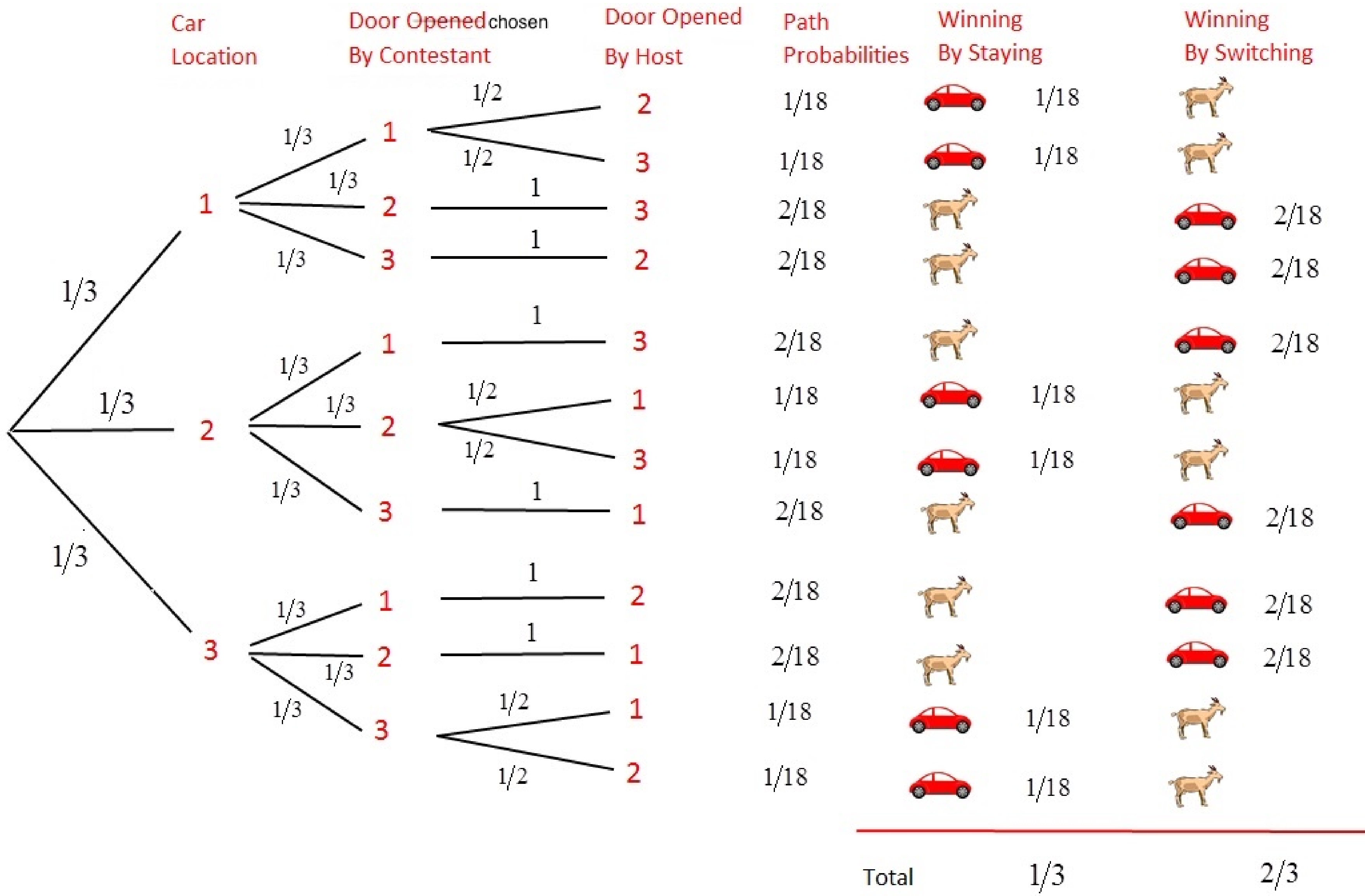
Step 3: Compute $P(C_1|H_2)$.

$$P(C_1|H_2) = \frac{P(H_2|C_1)P(C_1)}{P(H_2)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Step 4: Compute $P(C_3|H_2)$.

$$P(C_3|H_2) = \frac{P(H_2|C_3)P(C_3)}{P(H_2)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Thus, staying with Door 1 gives a $\frac{1}{3}$ chance of winning, while switching to Door 3 gives a $\frac{2}{3}$ chance. Tree diagram of probability path is as follows:


Car Location
Door Opened chosen By Contestant
Door Opened By Host
Path Probabilities
Winning By Staying
Winning By Switching
1/3
1/3
1/3
1
2
3
1/3
1/3
1/3
1
2
3
1/2
1/2
1
1
2
3
3
2
1/18
1/18
2/18
2/18
1/18
1/18
2/18
2/18
1
2
3
1
1/2
1/2
1
3
1
3
1
2/18
1/18
1/18
2/18
2/18
1/18
1/18
2/18
1
2
3
1
1
1/2
1/2
2
1
1
2
2/18
2/18
1/18
1/18
2/18
2/18
1/18
1/18
Total
1/3
2/3

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

Disclaimer