1 Exclusivity and Independence

1.1 Conditional Probability

Probabilities can change given more information. When we are given new information, we can restrict our sample space to outcomes that are consistent with this new information and update the probabilities in the following way.

Definition 1 (Conditional Probability). The *conditional probability* of A given B is, provided that $\mathbb{P}(B) > 0$, is

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

Remark 1. In general, $\mathbb{P}(A \mid B)$ and $\mathbb{P}(B \mid A)$ do not give you the same value, so the order matters.

Essentially, if we are given that B occurred, then we restrict the sample space Ω to the new sample space of possible outcomes B and recompute the probabilities. The new events are now $A \cap B$ (since events have to be subsets of B) and the normalization by $\mathbb{P}(B)$ ensures that conditional probabilities are probabilities (since the probability of B given B must be 1). Conditional probabilities are probabilities, they satisfy many of the same properties:

1. $0 \leq \mathbb{P}(A \mid B) \leq 1$
2. If A_1 and A_2 are disjoint, then $\mathbb{P}(A_1 \cup A_2 \mid B) = \mathbb{P}(A_1 \mid B) + \mathbb{P}(A_2 \mid B)$
3. $\mathbb{P}(A^c \mid B) = 1 - \mathbb{P}(A \mid B)$
4. $\mathbb{P}(\Omega \mid B) = 1 = \mathbb{P}(B \mid B)$

1.2 Mutually Independent and Mutually Exclusive Events

We extend the concept of the sum and product rule from counting to probabilities that are not necessarily uniform.

1.2.1 (Mutually) Exclusive Events

Recall that two events are *mutually exclusive* if they cannot happen at the same time.

Definition 2 (Exclusive Events). Two events A and B are *exclusive* if $A \cap B = \emptyset$. A sequence of events $A_1, A_2, ..., A_n$ are said to be *mutually exclusive* if

$$A_{i_1} \cap A_{i_2} = \emptyset,$$

for all possible $i_1 \neq i_2$.

Remark 2. Mutually exclusive events satisfy the analogue of the "sum" rule. If $A_1, A_2, \ldots, A_n$ are mutually exclusive events, then

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

This property is very useful because it mirrors how volumes are computed, in the sense that the volume of non-overlapping objects is equal to the sum of the volumes of individual objects.

1.2.2 (Mutually) Independent Events

Informally, two events are *independent* if they do not have an influence on each other.

Definition 3 (Independent Events I)**.** Two events A and B are independent, if

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

provided $\mathbb{P}(B) > 0$. (Or equivalently, $\mathbb{P}(B \mid A) = \mathbb{P}(B)$ provided that $\mathbb{P}(A) > 0$.)

From the definition of conditional probability, we arrive at the following equivalent notion.

Definition 4 (Independent Events II)**.** Two events A and B are said to be *independent* if

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

A sequence of events $A_1, A_2, \ldots, A_n$ are said to be *mutually independent* if

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\,\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}),$$

for all possible $1 \leq k \leq n$ and $1 \leq i_1 < \ldots < i_k \leq n$. Events that are *not independent* are *dependent*.

Remark 3. Mutually independent events satisfy the analogue of the "product" rule. If $A_1, A_2, \ldots, A_n$ are mutually independent events, then

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

This property is very useful because it mirrors how volumes are computed, in the sense that the volume of a high dimensional objects is equal to the product of it's one dimensional side lengths.

1.2.3 Mutually Exclusive vs Mutually Independent

It is important to not confuse mutually independent events with mutually exclusive events. They are actually completely "opposite" features, since if A and B are independent then the occurrence of B does not affect the occurrence of A

$$A, B \text{ independent} \iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) \iff \mathbb{P}(A \mid B) = \mathbb{P}(A)$$

but if A and B are mutually exclusive, then the occurrence of B means that A cannot occur

$$A, B \text{ mutually exclusive} \implies \mathbb{P}(A \cap B) = 0 \iff \mathbb{P}(A \mid B) = 0.$$

In fact, events A and B cannot be both mutually exclusive and independent (unless one of the events has probability zero (see Problem 1.11)).

1.3 Probability Rules

1.3.1 Inclusion Exclusion Principle:

We can generalize the sum rule to events that are not mutually exclusive.

Theorem 1 (*Inclusion–Exclusion Principle*)

For any events, A and B

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

General Form: *For arbitrary events* $A_1, A_2, \dots, A_n$ *with* $n \geq 2$,

$$\begin{aligned}\mathbb{P}(\cup_{i=1}^n A_i) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) \\ - \sum_{i<j<k<l} \mathbb{P}(A_i \cap A_j \cap A_k \cap A_l) + \cdots + (-1)^{n-1} \sum_{i<\cdots<n} \mathbb{P}(\cap_{i=1}^n A_i).\end{aligned}$$

The inclusion exclusion principle is tricky to compute when n is large, so we sometimes use the union bound to get an inequality of the union of events.

Corollary 1 (*Union Bound*)

For any events A *and* B,

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

General Form: *For arbitrary events* $A_1, A_2, \dots, A_n$ *with* $n \geq 2$,

$$\mathbb{P}(\cup_{i=1}^n A_i) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) = \sum_i \mathbb{P}(A_i).$$

1.3.2 Chain Rule

We can generalize the multiplication rule to events that are not mutually independent.

Theorem 2 (*Chain Rule*)

For any events A *and* B,

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

General Form: *For arbitrary events* $A_1, A_2, \dots, A_n$ *with* $n \geq 2$,

$$\begin{aligned}\mathbb{P}(\cap_{i=1}^n A_i) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \\ &= \prod_{k=1}^n \mathbb{P}(A_k \mid \cap_{j=1}^{k-1} A_j).\end{aligned}$$

1.3.3 Law of Total Probability

It is often easier to compute probabilities of events by breaking into cases.

Definition 5. A sequence of sets $B_1, B_2, ..., B_n$ are said to *partition* the sample space Ω if $B_i \cap B_j = \emptyset$ for all $i \neq j$, and $\cup_{j=1}^n B_j = \Omega$.

Essentially, breaking into the "cases" $B_1, \dots, B_k$ is valid whenever they form a partition because the condition $B_i \cap B_j = \emptyset$ ensures that our cases are distinct (so we don't overcount) and the $\cup_{j=1}^n B_j = \Omega$ condition ensures that we have not missed any cases (so we don't undercount). The simplest partition of Ω is B and B^c, which essentially breaks into the cases B happens and B does not happen.

Theorem 3 (*Law of total probability*)

For any events A *and* B,

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A \mid B)\, \mathbb{P}(B) + \mathbb{P}(A \mid B^c)\, \mathbb{P}(B^c)$$

General Form: *Suppose that $B_1, B_2, ..., B_n$ partition of Ω. Then for any event A,*

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

1.3.4 Bayes' Theorem

We can "flip" the terms in event and conditioning event using Bayes Theorem.

Theorem 4 (*Bayes' Theorem*)

For any events A and B,

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\,\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B)\,\mathbb{P}(B)}{\mathbb{P}(A \mid B)\,\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\,\mathbb{P}(B^c)}$$

General Form: *Suppose that $B_1, B_2, ..., B_n$ partition S. Then for any event A,*

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A \mid B_j)\,\mathbb{P}(B_j)}.$$

In Bayesian statistics, we often call $\mathbb{P}(B)$ the *prior* and $\mathbb{P}(B \mid A)$ the *posterior*. Bayes' theorem gives us a rule for how our prior probability adapts given more information A.

1.4 Example Problems

Problem 1.1. You roll a die and your friend looks at what number you rolled.

- What's the probability that it's a 6?
- What's the probability that it's a 6 given that your friend tells you that it's an even number?
- What's the probability that it's a 6 given that your friend tells you that it's an odd number?

Solution 1.1.

1. Since all outcomes are equally likely,

$$\mathbb{P}(\text{roll } 6) = \frac{1}{6}.$$

Remark 4. This answer makes sense since in the absence of more information, the sample space of possible outcomes is still $\{1, 2, 3, 4, 5, 6\}$, and the probability is uniform over this set.

2. By definition

$$\mathbb{P}(\text{roll } 6 \mid \text{even}) = \frac{\mathbb{P}(\text{roll } 6, \text{even})}{\mathbb{P}(\text{even})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Remark 5. This answer makes sense since if we know that the roll was even, the restricted sample space of possible outcomes is now $\{2, 4, 6\}$, and we have no reason to expect that any of these possible outcomes is more likely than the other, so the probability is now $1/3$.

3. By definition
$$\mathbb{P}(\text{roll } 6 \mid \text{odd}) = \frac{\mathbb{P}(\text{roll } 6, \text{odd})}{\mathbb{P}(\text{odd})} = \frac{0}{\frac{1}{2}} = 0.$$

Remark 6. This answer makes sense since if we know that the roll was odd, the restricted sample space of possible outcomes is now $\{1, 3, 5\}$. Since 6 is no longer a possible outcome, the probability is now 0.

Problem 1.2. Consider rolling two fair six sided dice, and let
$$A = \{\text{the sum is } 10\}, \quad B = \{\text{the first die is a } 6\}, \quad C = \{\text{the sum is } 7\}.$$

1. Compute $\mathbb{P}(A \mid B)$
2. Compute $\mathbb{P}(B \mid A)$
3. Compute $\mathbb{P}(A \mid C)$
4. Compute $\mathbb{P}(C \mid B)$

Solution 1.2. This problem is a direct application of the definition of conditional probabilities,

1. Since $\mathbb{P}(A \cap B) = \mathbb{P}((6, 4)) = 1/36$ we find $\mathbb{P}(A \mid B) = \frac{1/36}{1/6} = \frac{1}{6}$.
2. Similarly, $\mathbb{P}(B \mid A) = \frac{1/36}{1/12} = \frac{1}{3}$.
3. Since $A \cap C = \emptyset$, we have $\mathbb{P}(A \cap C) = 0$ and $\mathbb{P}(A \mid C) = \frac{0}{\mathbb{P}(C)} = 0$.
4. Since $\mathbb{P}(C \cap B) = \mathbb{P}((6, 1)) = 1/36$ we find $\mathbb{P}(C \mid B) = \frac{1/36}{1/6} = \frac{1}{6}$.

Problem 1.3. Consider rolling two fair six sided dice, and let
$$A = \{\text{the sum is } 10\}, \quad B = \{\text{the first die is a } 6\}, \quad C = \{\text{the sum is } 7\}.$$

1. Are A and B independent?
2. Are A and C independent?
3. Are B and C independent?

Solution 1.3. We compute all relevant probabilities:
$$\begin{aligned}
&\mathbb{P}(A) = \mathbb{P}(\{(5, 5), (6, 4), (4, 6)\}) = 3/36 = 1/12 \\
&\mathbb{P}(B) = 1/6 \\
&\mathbb{P}(C) = 6/36 = 1/6.
\end{aligned}$$

We now check the definition of independence
$$\begin{aligned}
&\mathbb{P}(A \cap B) = \mathbb{P}((6, 4)) = 1/36 \neq \mathbb{P}(A)\,\mathbb{P}(B) && \Rightarrow A \text{ and } B \text{ are not independent} \\
&\mathbb{P}(A \cap C) = 0 \neq \mathbb{P}(A)\,\mathbb{P}(C) && \Rightarrow A \text{ and } C \text{ not independent} \\
&\mathbb{P}(B \cap C) = \mathbb{P}((6, 1)) = 1/36 = \mathbb{P}(B)\,\mathbb{P}(C) && \Rightarrow B \text{ and } C \text{ independent}
\end{aligned}$$

Problem 1.4. A survey company found that 63% of Canadians support Canadian Tire and 80% support Tim Hortons, while 51% of Canadians support both Tim Hortons and Canadian Tire. Let A denote the event that an individual supports Canadian Tire and B the event that an individual supports Tim Hortons.

1. Suppose a Canadian is selected at random, what is the probability that the individual supports Canadian Tire or Tim Hortons?
2. Suppose a Canadian is selected at random, what is the probability that the individual supports Tim Hortons but not Canadian Tire?

Solution 1.4.

Part 1: We are given that $\mathbb{P}(A) = 0.63$, $\mathbb{P}(B) = 0.8$ and $\mathbb{P}(A \cap B) = 0.51$ and we are asked to find $\mathbb{P}(A \cup B)$. By the inclusion exclusion principle,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.63 + 0.8 - 0.51 = 0.92.$$

Part 2: We are given that $\mathbb{P}(A) = 0.63$, $\mathbb{P}(B) = 0.8$ and $\mathbb{P}(A \cap B) = 0.51$ and we are asked to find $\mathbb{P}(A^c \cap B)$. By the law of total probability,

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) \implies \mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.8 - 0.51 = 0.29.$$

Problem 1.5. The probability a randomly selected male is colour-blind is 0.05, whereas the probability a female is colour-blind is only 0.0025. If the population is 45% male, what is the fraction that is colour-blind?

Solution 1.5. Let

$$C = \text{“randomly selected person is colorblind”} \quad \text{and} \quad M = \text{“randomly selected person is male.”}$$

We are given

$$\mathbb{P}(C \mid M) = 0.05; \;\; \mathbb{P}(C \mid M^c) = 0.0025; \;\; \mathbb{P}(M) = 0.45; \;\; \mathbb{P}(M^c) = 0.55$$

and want to find $\mathbb{P}(C)$. The law of total probability implies that

$$\mathbb{P}(C) = \mathbb{P}(C \mid M)P(M) + \mathbb{P}(C \mid M^c)P(M^c) = 0.05 \cdot 0.45 + 0.0025 \cdot 0.55 = 0.023875.$$

Problem 1.6. Justin has 12 red hats and 7 green hats. On Monday, Justin picks hats for his Monday and Wednesday lectures. He picks one hat at random (for the Monday lecture) and then, without replacement, another hat (for the Wednesday lecture).

1. What is the probability that both hats are red?
2. What is the probability that Justin wears a red hat on Wednesday?

Solution 1.6. The second hat I pick depends on the outcome of the first hat, so we should use conditional probabilities to compute our answers.

Part 1. Let

$$R_1 = \text{ “first pair is red”} \quad \text{and } R_2 = \text{ “second pair is red”}.$$

By the chain rule,

$$\begin{aligned}\mathbb{P}(R_1 \cap R_2) &= \mathbb{P}(R_1) \cdot \mathbb{P}(R_2 \mid R_1) \\ &= \frac{12}{19} \cdot \frac{11}{18} = \frac{22}{57} \approx 0.386\end{aligned}$$

Part 2: By the law of total probability,

$$\begin{aligned}\mathbb{P}(R_2) &= \mathbb{P}(R_2 \cap R_1) + \mathbb{P}(R_2 \cap R_1^c) \\ &= \mathbb{P}(R_2 \mid R_1)\,\mathbb{P}(R_1) + \mathbb{P}(R_2 \mid R_1^c)\,\mathbb{P}(R_1^c) \\ &= \frac{12}{19} \cdot \frac{11}{18} + \frac{7}{19} \cdot \frac{12}{18} \\ &= \frac{12}{19} \approx 0.632\end{aligned}$$

Problem 1.7. If I pick a red hat, there is a 20% chance somebody will laugh. If I pick a green hat, there is a 10% chance somebody will laugh. 70% of the time I choose a red hat, and 30% of the time a green hat. Given that somebody laughed, what is the probability that I picked a green hat?

Solution 1.7. Denote by

$$L = \text{“somebody laughed”} \quad \text{and} \quad G = \text{“I wore a green hat”}$$

We are given that $\mathbb{P}(L \mid R) = 0.2$, $\mathbb{P}(L \mid G) = 0.1$, $\mathbb{P}(R) = 0.7$ and $\mathbb{P}(G) = 0.3$. We need to compute the “flipped” conditional probability $\mathbb{P}(G \mid L)$. By Bayes' rule

$$\mathbb{P}(G \mid L) = \frac{\mathbb{P}(L \mid G)\,\mathbb{P}(G)}{\mathbb{P}(L \mid R)\,\mathbb{P}(R) + \mathbb{P}(L \mid G)\,\mathbb{P}(G)} = \frac{0.1 \cdot 0.3}{0.2 \cdot 0.7 + 0.1 \cdot 0.3} = 0.1765.$$

Problem 1.8. In an insurance portfolio 10% of the policy holders are in Class A1 (high risk), 40% are in Class A2 (medium risk), and 50% are in Class A3 (low risk). The probability there is a claim on a Class A1 policy in a given year is 0.10; similar probabilities for Classes A2 and A3 are 0.05 and 0.02.

1. Find the probability of a claim.
2. Find the probability that if a claim is made, it is made on a Class A1 policy.

Solution 1.8. Let

$$A_j = \text{“customer is in class } j\text{”} \qquad \text{and} \qquad C = \text{“there's a claim”}.$$

We are given

$$\mathbb{P}(A_1) = 0.1,\ \mathbb{P}(A_2) = 0.4,\ \mathbb{P}(A_3) = 0.5$$

and

$$\mathbb{P}(C \mid A_1) = 0.1,\ \mathbb{P}(C \mid A_2) = 0.05,\ \mathbb{P}(C \mid A_3) = 0.02.$$

1. We want to find $\mathbb{P}(C)$. Since there are only 3 classes, A_1, A_2, A_3 form a partition, so the law of total probability implies that

$$\mathbb{P}(C) = \sum_{j=1}^{3} \mathbb{P}(C \mid A_j)\,\mathbb{P}(A_j) = 0.1 \cdot 0.1 + 0.05 \cdot 0.4 + 0.02 \cdot 0.5 = 0.04.$$

2. We want to find $\mathbb{P}(A_1 \mid C)$. By Bayes' rule

$$\mathbb{P}(A_1 \mid C) = \frac{\mathbb{P}(C \mid A_1)\,\mathbb{P}(A_1)}{\sum_{j=1}^{3} \mathbb{P}(C \mid A_j)\,\mathbb{P}(A_j)} = \frac{0.1 \cdot 0.1}{0.1 \cdot 0.1 + 0.05 \cdot 0.4 + 0.02 \cdot 0.5} = 0.25.$$

Problem 1.9. Motor vehicles sold to individuals are classified as either cars or light trucks (including SUVs) and as either domestic or imported. In a recent year among all vehicles, 69% of were light trucks, 78% were domestic, and 55% were domestic light trucks. Let A be the event that a randomly selected vehicle is a car, and B the event that it is imported.

1. Given that the vehicle is imported, what is the probability that it is a light truck?
2. Are the events "vehicle is a light truck" and "vehicle is domestic" independent? Are they mutually exclusive?
3. Motor vehicles can also be classified by their colour. Police officers tend to stop red-coloured vehicles more often than any other colour, explaining why 81% of vehicles have a different colour than red. Suppose that whether a vehicle is an imported car is independent of its color. Find the probability that a vehicle is red and has at least one of the following two characteristics: domestic or light truck.

Solution 1.9. We are given $\mathbb{P}(A^c) = 0.69$, $\mathbb{P}(B^c) = 0.78$, and $\mathbb{P}(A^c \cap B^c) = 0.55$. This implies that

$$\mathbb{P}(A) = 1 - 0.69 = 0.31 \qquad \mathbb{P}(B) = 1 - 0.78 = 0.22.$$

Furthermore, by the inclusion exclusion principle,

$$\mathbb{P}(A^c \cup B^c) = \mathbb{P}(A^c) + \mathbb{P}(B^c) - \mathbb{P}(A^c \cap B^c) = 0.69 + 0.78 - 0.55 = 0.92.$$

Therefore, DeMorgan's law implies that $(A^c \cup B^c)^c = A \cap B$, so

$$\mathbb{P}(A \cap B) = \mathbb{P}((A^c \cup B^c)^c) = 1 - \mathbb{P}(A^c \cup B^c) = 1 - 0.92 = 0.08.$$

1. We want to find $\mathbb{P}(A^c \mid B)$. By the definition of conditional probability

$$\mathbb{P}(A^c \mid B) = \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B) - \mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0.22 - 0.08}{0.22} = 0.6364.$$

Alternative Solution: By the definition of conditional probability

$$\mathbb{P}(A^c \mid B) = \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A^c) - \mathbb{P}(A^c \cap B^c)}{\mathbb{P}(B)} = \frac{0.69 - 0.55}{0.22} = 0.6364.$$

2. We can check the definition directly,
   - Since $\mathbb{P}(A^c \cap B^c) = 0.55 \neq 0.69 \cdot 0.78 = \mathbb{P}(A^c)\,\mathbb{P}(B^c)$, the events are not independent.
   - Since $\mathbb{P}(A^c \cap B^c) = 0.55 \neq 0$, the events are not mutually exclusive, because if A^c and B^c were exclusive, it must be the case that $\mathbb{P}(A^c \cap B^c) = 0$.
3. Let C be the event that a vehicle is red. We want to find $\mathbb{P}(C \cap (A^c \cup B^c))$,

$$\begin{aligned}
\mathbb{P}(C \cap (A^c \cup B^c)) &= \mathbb{P}(C \cap (A \cap B)^c), && \text{using De Morgan's law} \\
&= \mathbb{P}(C)\,\mathbb{P}[(A \cap B)^c], && \text{due to independence of } A \cap B \text{ and } C \\
&= (1 - 0.81)(1 - 0.08) \\
&= 0.175.
\end{aligned}$$

Alternative Solution: Let C be the event that a vehicle is red. We want to find $\mathbb{P}(C\cap(A^c\cup B^c))$,

$$\begin{aligned}
\mathbb{P}(C\cap(A^c\cup B^c)) &= \mathbb{P}(C\cap(A\cap B)^c), \quad \text{using De Morgan's law} \\
&= \mathbb{P}(C)\,\mathbb{P}[(A\cap B)^c], \quad \text{due to independence of } A\cap B \text{ and } C \\
&= \mathbb{P}(C)\,\mathbb{P}[A^c\cup B^c], \quad \text{using De Morgan's law} \\
&= (1-0.81)\cdot 0.92 \\
&= 0.175.
\end{aligned}$$

Problem 1.10. You have n identical looking keys on a chain, and one opens your office door. Suppose you try the keys in a random order (without selecting an incorrect key more than once). What is the probability the k'th key opens the door?

Solution 1.10. Let A_i denote the event that the ith key opens the door. We want to find

$$\mathbb{P}(A_1^c\cap A_2^c\cap\cdots\cap A_{k-1}^c\cap A_k).$$

This problem is computable using the chain rule,

$$\begin{aligned}
\mathbb{P}(A_1^c\cap A_2^c\cap\cdots\cap A_{k-1}^c\cap A_k) &= \mathbb{P}(A_1^c)\cdot\mathbb{P}(A_2^c\mid A_1^c)\ldots\mathbb{P}(A_k\mid(A_1^c\cap A_2^c\cap\cdots\cap A_{k-1}^c) \\
&= \frac{n-1}{n}\cdot\frac{n-2}{n-1}\cdots\frac{n-k+1}{n-k+2}\cdot\frac{1}{n-k+1}=\frac{1}{n}.
\end{aligned}$$

We used the fact that

1. $\mathbb{P}(A_1^c)=\frac{n-1}{n}$ because there are $n-1$ wrong keys out of n keys in total,
2. $\mathbb{P}(A_2^c\mid A_1^c)=\frac{n-2}{n-1}$ since there are $n-2$ wrong keys out of the remaining $n-1$ keys, because we know that the first key we tried failed
3. $\mathbb{P}(A_j^c\mid A_1^c\cap\cdots\cap A_{j-1}^c)=\frac{n-j}{n-j+1}$ by inductive reasoning
4. and lastly, $\mathbb{P}(A_k\mid(A_1^c\cap A_2^c\cap\cdots\cap A_{k-1}^c)=\frac{1}{n-k+1}$ since there is one correct key out of the remaining $n-(k-1)$ keys (because we tried $k-1$ keys that failed to open the lock).

Alternative Solution I: The sample space for this problem are the ordered tuples of length n. If we know that the right key is used on the kth try, there are $(n-1)!$ choices for the other keys. So

$$\mathbb{P}(A_1^c\cap A_2^c\cap\cdots\cap A_{k-1}^c\cap A_k)=\frac{(n-1)!}{n!}=\frac{1}{n}.$$

Alternative Solution II: We can take the sample space to be the location of the right key out of n possible locations. Therefore, $\Omega=[n]$ and the location of the right key is uniform over this sample space by symmetry, so

$$\mathbb{P}(A_1^c\cap A_2^c\cap\cdots\cap A_{k-1}^c\cap A_k)=\frac{1}{|\Omega|}=\frac{1}{n}.$$

1.5 Proofs of Key Results

Problem 1.11. Suppose that A and B are independent and mutually exclusive. Show that $\mathbb{P}(A)=0$ or $\mathbb{P}(B)=0$.

Solution 1.11. On one hand, if A and B are independent, then

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

On the other hand, if A and B are exclusive

$$\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0.$$

For both to hold simultaneously, we must have $\mathbb{P}(A)\,\mathbb{P}(B) = 0$, so either $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$.

Problem 1.12. Prove the inclusion exclusion principle (Theorem 1).

Solution 1.12. Notice that A and $(B \setminus A)$ are exclusive events such that $A \cup (B \setminus A) = A \cup B$, so by countable additvity

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$$

Next, notice that $B \setminus A$ and $A \cap B$ are exclusive and $(A \cap B) \cup (B \setminus A) = A \cap B$,

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) \implies \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

which implies our result. The general statement follows from a straightforward but tedious induction argument. We have

$$\begin{aligned}\mathbb{P}(\cup_{i=1}^{n} A_i) = \mathbb{P}((\cup_{i=1}^{n-1} A_i) \cup A_n) &= \mathbb{P}(\cup_{i=1}^{n-1} A_i) + \mathbb{P}(A_n) - \mathbb{P}((\cup_{i=1}^{n-1} A_i) \cap A_n) \\ &= \mathbb{P}(\cup_{i=1}^{n-1} A_i) + \mathbb{P}(A_n) - \mathbb{P}(\cup_{i=1}^{n-1} (A_i \cap A_n))\end{aligned}$$

The formula is now expressed as the union of up to $n-1$ sets, so we can simplify using the inductive hypothesis to get our result.

Problem 1.13. Prove the chain rule for conditional probabilities (Theorem 2).

Solution 1.13. By the definition of conditional probabilities,

$$\begin{aligned}\mathbb{P}(A_{n-1} \mid A_1 \cap \cdots \cap A_{n-2})\,\mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}) &= \frac{\mathbb{P}(A_1 \cap \cdots \cap A_{n-1})}{\mathbb{P}(A_1 \cap \cdots \cap A_{n-2})} \cdot \frac{\mathbb{P}(A_1 \cap \cdots \cap A_n)}{\mathbb{P}(A_1 \cap \cdots \cap A_{n-1})} \\ &= \mathbb{P}(A_n \cap A_{n-1} \mid A_1 \cap \cdots \cap A_{n-2}).\end{aligned}$$

The exact same computation implies that

$$\mathbb{P}(A_n \cap A_{n-1} \mid A_1 \cap \cdots \cap A_{n-2})\,\mathbb{P}(A_{n-2} \mid A_1 \cap \cdots \cap A_{n-3}) = \mathbb{P}(A_n \cap A_{n-1} \cap A_{n-2} \mid A_1 \cap \cdots \cap A_{n-3}).$$

Each time we do this computation, we can move the right most set on the right of the $\mid$ to the left, so continuing inductively by computing the right most products implies our result.

Problem 1.14. Prove the law of total probability (Theorem 3).

Solution 1.14. Since $B_1, \dots B_k$ partition the sample space, we have

$$A = A \cap S = A \cap (\cup_{j=1}^{n} B_j) = \cup_{j=1}^{n} (A \cap B_j)$$

and the events $\{(A \cap B_1), \dots, (A \cap B_n)\}$ are mutually exclusive because the B_j are. Therefore, by countable additivity,

$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \cdots + \mathbb{P}(A \cap B_n).$$

To prove the second equality, we can use the the chain rule to conclude that $\mathbb{P}(A \cap B_j) = \mathbb{P}(A|B_j)\,\mathbb{P}(B_j)$ for all j, so

$$\mathbb{P}(A) = \mathbb{P}(A \mid B_1)\,\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2) P(B_2) + \cdots + \mathbb{P}(A \mid B_n)\,\mathbb{P}(B_n).$$

Problem 1.15. Prove Bayes rule (Theorem 4).

Solution 1.15. We start from the definition of conditional probability,

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

By the chain rule, $\mathbb{P}(B_i \cap A) = \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)$, so

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

To get the second equality, the law of total probability states that

$$\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}(A \mid B_j)\,\mathbb{P}(B_j)$$

Problem 1.16. Suppose that $A \subseteq B$. Show that $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Solution 1.16. This property is called *monotonicity*. By the law of total probability,

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

Since $A \subseteq B$, we have $\mathbb{P}(A \cap B) = \mathbb{P}(A)$ and $\mathbb{P}(A^c \cap B) \geq 0$, so

$$\mathbb{P}(A) \leq \mathbb{P}(B).$$

Problem 1.17. Show that for arbitrary events $A_1, A_2, \ldots, A_n$ (Corollary 1),

$$\mathbb{P}(\cup_{k=1}^{n} A_k) \leq \sum_{k=1}^{n} \mathbb{P}(A_k).$$

Solution 1.17. This is called the *union bound*. By the inclusion exclusion principle, we see that

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

since $\mathbb{P}(A \cap B) \geq 0$. The general statement then follows by (strong) induction. We have

$$\begin{aligned}
\mathbb{P}(A_1 \cup A_2 \cup A_{n-1} \cup A_n) &= \mathbb{P}((A_1 \cup A_2 \cup A_{n-1}) \cup A_n) \\
&\leq \mathbb{P}(A_1 \cup A_2 \cup A_{n-1}) + \mathbb{P}(A_n) \\
&\leq \sum_{k=1}^{n-1} \mathbb{P}(A_k) + \mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}(A_k).
\end{aligned}$$

Remark 7. It is true that equality is attained if $A_1, \ldots, A_n$ are mutually exclusive. However, we can have equality even if $A_1, \ldots, A_n$ are not mutually exclusive. We can consider the distribution on the two points $\{a, b\}$ such that $\mathbb{P}(\{a\}) = 0$ and $\mathbb{P}(\{b\}) = 1$. Then $A = \{a, b\}$ and $B = \{a\}$ satisfy the union bound, but they are not mutually exclusive.