

Exercise Walkthrough: The Inclusion-Exclusion Principle

Justin Lanfermann

25. June 2025

1 Introduction

This document provides a detailed, step-by-step walkthrough for proving the **Principle of Inclusion-Exclusion**. This principle is a cornerstone of combinatorics and probability theory. It provides a systematic way to find the probability of the union of multiple events, which is essential when events are not mutually exclusive.

The proof will be structured using the method of **complete induction** [1]. We will first establish a base case and then show that if the formula holds for n events, it must also hold for $n + 1$ events. Throughout the proof, we will refer to definitions and properties from the "Discrete Probability Theory" script by Niki Kilbertus.

Exercise 1 (Inclusion-Exclusion Formula). *Let (Ω, \mathcal{A}, P) be a probability space and let A_1, A_2, \dots, A_n be a finite collection of events in \mathcal{A} . Prove that:*

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k})$$

2 The Proof by Induction

2.1 Step 1: The Base Case (n=2)

Goal First, we need to show that the formula holds for the simplest non-trivial case, which is $n = 2$. We must prove that $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

Reasoning and Steps The core idea is to decompose the union $A_1 \cup A_2$ into disjoint sets. By doing so, we can use the additivity property of a probability measure [2].

1. **Decomposition into disjoint sets:** We can write the union $A_1 \cup A_2$ as the union of two disjoint sets: A_2 and the part of A_1 that is not in A_2 , which is $A_1 \setminus A_2$.

$$A_1 \cup A_2 = (A_1 \setminus A_2) \cup A_2$$

Since an element in $A_1 \setminus A_2$ is by definition not in A_2 , these two sets are disjoint.

2. **Apply Additivity:** Because the sets are disjoint, we can apply the additivity property of the probability measure P (from *Definition 1.18* in the script):

$$P(A_1 \cup A_2) = P(A_1 \setminus A_2) + P(A_2)$$

3. **Use the Difference Rule:** Now we need to handle the term $P(A_1 \setminus A_2)$. From *Proposition 1.17 (ii)* in the script, we know the formula for the probability of a set difference [3]. This rule states that if $B \subseteq A$, then $P(A \setminus B) = P(A) - P(B)$. In our case, $A_1 \cap A_2$ is a subset of A_1 . Therefore:

$$P(A_1 \setminus A_2) = P(A_1) - P(A_1 \cap A_2)$$

4. **Combine:** Substituting this back into our equation from step 2, we get:

$$P(A_1 \cup A_2) = (P(A_1) - P(A_1 \cap A_2)) + P(A_2)$$

Rearranging the terms gives the desired result for the base case:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

The base case holds.

2.2 Step 2: The Inductive Hypothesis (IH)

Goal We assume that the inclusion-exclusion formula is true for some arbitrary integer $n \geq 2$.

Assumption Let's assume for any collection of n events B_1, \dots, B_n , the following holds:

$$P\left(\bigcup_{i=1}^n B_i\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(B_{i_1} \cap \dots \cap B_{i_k})$$

2.3 Step 3: The Inductive Step (IS)

Goal Now, we must prove that the formula also holds for $n + 1$ events, using our base case and the inductive hypothesis. We want to show:

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = \sum_{k=1}^{n+1} (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n+1} P(A_{i_1} \cap \dots \cap A_{i_k})$$

Reasoning and Steps The strategy is to treat the union of $n + 1$ sets as a union of two "meta-sets" and apply our proven base case. This will create terms where we can apply our inductive hypothesis.

1. **Group the terms:** Let's group the first n events together. Let $B = \bigcup_{i=1}^n A_i$. Then our union becomes a union of two sets:

$$\bigcup_{i=1}^{n+1} A_i = \left(\bigcup_{i=1}^n A_i\right) \cup A_{n+1} = B \cup A_{n+1}$$

2. **Apply the Base Case:** We can now apply our proven formula for $n = 2$ to B and A_{n+1} :

$$P(B \cup A_{n+1}) = P(B) + P(A_{n+1}) - P(B \cap A_{n+1}) \quad (1)$$

3. **Expand the terms using IH:** We have two complex terms, $P(B)$ and $P(B \cap A_{n+1})$, that we need to expand.

For $P(B) = P(\bigcup_{i=1}^n A_i)$, we can directly apply the Inductive Hypothesis (IH):

$$P(B) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \quad (2)$$

For $P(B \cap A_{n+1})$, we first use the distributive law for sets [4]:

$$B \cap A_{n+1} = \left(\bigcup_{i=1}^n A_i\right) \cap A_{n+1} = \bigcup_{i=1}^n (A_i \cap A_{n+1})$$

This is a union of n sets, $(A_1 \cap A_{n+1}), \dots, (A_n \cap A_{n+1})$. We can apply our IH to this union:

$$\begin{aligned}
P(B \cap A_{n+1}) &= P\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right) \\
&= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P((A_{i_1} \cap A_{n+1}) \cap \dots \cap (A_{i_k} \cap A_{n+1})) \\
&= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k} \cap A_{n+1}) \tag{3}
\end{aligned}$$

4. **Combine and Simplify:** Now substitute (2) and (3) back into (1):

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n+1} A_i\right) &= \underbrace{\left(\sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(\cap_{j=1}^k A_{i_j})\right)}_{P(B)} + P(A_{n+1}) \\
&\quad - \underbrace{\left(\sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(\cap_{j=1}^k A_{i_j} \cap A_{n+1})\right)}_{P(B \cap A_{n+1})}
\end{aligned}$$

Let's distribute the minus sign in the last term and rearrange:

$$\begin{aligned}
&= P(A_{n+1}) + \sum_{k=1}^n (-1)^{k-1} \sum_{I \subseteq [n], |I|=k} P\left(\bigcap_{i \in I} A_i\right) \\
&\quad + \sum_{k=1}^n (-1)^k \sum_{J \subseteq [n], |J|=k} P\left(\left(\bigcap_{j \in J} A_j\right) \cap A_{n+1}\right)
\end{aligned}$$

Here we've used set notation for the indices for clarity. Let's analyze the sums. The target formula for $n+1$ involves sums over all subsets of $[n+1]$. Any subset of $[n+1]$ of size k either:

- (a) Does **not** contain the element $n+1$. These are exactly the subsets of $[n]$ of size k .
- (b) **Does** contain the element $n+1$. These are formed by taking a subset of $[n]$ of size $k-1$ and adding $n+1$.

Our expression perfectly captures this structure [5]:

- The term $P(A_{n+1})$ is the sum for intersections of size $k=1$ that only contains A_{n+1} .
- The first sum $\sum_{k=1}^n \dots$ covers all intersections of events from $\{A_1, \dots, A_n\}$.
- The second sum $\sum_{k=1}^n \dots$ covers all intersections that are guaranteed to include A_{n+1} .

Let's combine them term by term for each intersection size k' from 1 to $n+1$:

- For $k'=1$: We get $\sum_{i=1}^n P(A_i)$ from the first sum and $P(A_{n+1})$ from the standalone term. Total: $\sum_{i=1}^{n+1} P(A_i)$. This matches.
- For $k' \in \{2, \dots, n\}$: The term is $(-1)^{k'-1}$. Intersections of size k' come from two places:

- (a) Those not containing A_{n+1} (from the first sum, with $k = k'$).
- (b) Those containing A_{n+1} (from the second sum, with $k = k' - 1$).

Combining these gives precisely the sum over all intersections of size k' from $\{A_1, \dots, A_{n+1}\}$.

- For $k' = n + 1$: This term only comes from the second sum, when $k = n$. The term is $(-1)^n P(A_1 \cap \dots \cap A_n \cap A_{n+1})$. This is the correct final term for the $n + 1$ formula.

Thus, by rearranging and grouping the summations, we arrive at the formula for $n + 1$:

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = \sum_{k=1}^{n+1} (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n+1} P(A_{i_1} \cap \dots \cap A_{i_k})$$

This completes the inductive step.

3 Conclusion

We have successfully shown that:

1. The formula holds for the base case $n = 2$.
2. If the formula holds for any n events, it also holds for $n + 1$ events.

By the principle of mathematical induction, the Inclusion-Exclusion formula is true for any finite number of events $n \geq 2$.

Check Your Understanding

To solidify your understanding, try writing out the formula for $n = 3$ events, $P(A \cup B \cup C)$. You should get:

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

Does this match the general formula we just proved?

In-depth Explanations

[1] **The Principle of Mathematical Induction** Induction is a powerful proof technique used to prove statements about all natural numbers (or any well-ordered set). It works like a chain of dominoes:

- **Base Case:** You prove the statement is true for the first number (e.g., $n = 1$ or $n = 2$). This is like tipping the first domino.
- **Inductive Step:** You prove that *if* the statement is true for an arbitrary number n (this is the Inductive Hypothesis), *then* it must also be true for the next number, $n + 1$. This is like showing that any domino is close enough to the next one to knock it over.

If you can prove both parts, you have proven the statement for all numbers starting from your base case, because the first one falls, which makes the second one fall, which makes the third one fall, and so on, ad infinitum.

[2] **Additivity of Probability for Disjoint Sets** This is a direct consequence of the third axiom of a probability measure, σ -additivity, as defined in *Definition 1.16 (iii)* and specialized for probability in *Definition 1.18*. The axiom states that for any *countable* collection of pairwise disjoint sets $(A_i)_{i \in \mathbb{N}}$ in \mathcal{A} , we have:

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

For a finite number of disjoint sets, say A_1, A_2 , we can consider the sequence $A_1, A_2, \emptyset, \emptyset, \dots$. The axiom then implies $P(A_1 \cup A_2) = P(A_1) + P(A_2) + P(\emptyset) + \dots = P(A_1) + P(A_2)$, since $P(\emptyset) = 0$.

[3] **Probability of a Set Difference** The formula $P(A \setminus B) = P(A) - P(A \cap B)$ is a basic property of measures shown in *Proposition 1.17 (ii)* (though stated there for the case when $B \subseteq A$). To see it more generally, note that $A = (A \setminus B) \cup (A \cap B)$. These two sets are disjoint. By additivity [2], we have $P(A) = P(A \setminus B) + P(A \cap B)$. Rearranging this equation gives the desired result.

[4] **Distributive Law for Sets** Set operations follow distributive laws, similar to multiplication over addition in arithmetic. Specifically, intersection distributes over union:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

This can be extended to unions of multiple sets:

$$\left(\bigcup_{i=1}^n A_i\right) \cap B = \bigcup_{i=1}^n (A_i \cap B)$$

This is a standard result from set theory (see *Appendix A.1*).

[5] **Combinatorial Argument (Pascal's Identity)** The way we combined the sums in the inductive step is a probabilistic version of a famous combinatorial identity called Pascal's Identity: $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$. The sum $\sum_{1 \leq i_1 < \dots < i_k \leq n+1}$ is a sum over all $\binom{n+1}{k}$ subsets of size k from the set $\{1, \dots, n+1\}$. Our proof showed that the sum over all intersections of size k is formed by:

- The sum over the $\binom{n}{k}$ intersections that do not involve A_{n+1} .
- The sum over the $\binom{n}{k-1}$ intersections that are formed by taking $k-1$ events from the first n and intersecting them with A_{n+1} .

Adding these two collections of intersections gives all $\binom{n+1}{k}$ possible intersections of size k . Our proof carefully tracked the alternating signs to show they combined correctly as well.