

Exercise Walkthrough: Total Probability and Bayes' Rule

Justin Lanfermann

25. June 2025

1 Overview

This document provides a step-by-step solution to the exercise concerning conditional probability, the law of total probability, and Bayes' rule. The goal is not just to present the answer, but to explain the reasoning behind each step, grounding our arguments in the formal, measure-theoretic framework of probability established in the lecture script.

We will prove three fundamental results:

1. That the conditional probability map $\tilde{P}(A) := P(A|B)$ is itself a valid probability measure.
2. The Law of Total Probability, which allows us to calculate the probability of an event by conditioning on a partition of the sample space.
3. Bayes' Rule, a powerful formula for inverting conditional probabilities.

Throughout this walkthrough, we will reference definitions from the lecture script and use clickable notes like [\[1\]](#) for in-depth explanations of key concepts.

2 Part (i): Verifying the Probability Measure

Statement

Let (Ω, \mathcal{A}, P) be a probability space [\[1\]](#) and $B \in \mathcal{A}$ with $P(B) > 0$. Show that the map

$$\tilde{P} : \mathcal{A} \rightarrow [0, 1], \quad A \mapsto \tilde{P}(A) := P(A|B)$$

is a probability measure on (Ω, \mathcal{A}) .

Proof Walkthrough

To show that \tilde{P} is a probability measure, we must verify the four properties laid out in **Definition 1.18** (referencing Def. 1.16) from the script. A function $\mu : \mathcal{A} \rightarrow [0, \infty]$ is a probability measure if it satisfies:

1. $\mu(\emptyset) = 0$
2. $\mu(A) \geq 0$ for all $A \in \mathcal{A}$ (Non-negativity)
3. $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ for any countable collection of pairwise disjoint sets $A_i \in \mathcal{A}$ (σ -additivity) [\[3\]](#).
4. $\mu(\Omega) = 1$ (Normalization)

Let's check these properties for \tilde{P} one by one.

1. Probability of the Empty Set: We start by applying the definition of \tilde{P} to the empty set \emptyset .

$$\begin{aligned}
\tilde{P}(\emptyset) &:= P(\emptyset|B) && \text{(Definition of } \tilde{P}) \\
&= \frac{P(\emptyset \cap B)}{P(B)} && \text{(Definition of Conditional Probability [2], script Def. 1.64)} \\
&= \frac{P(\emptyset)}{P(B)} && \text{(Since } \emptyset \cap B = \emptyset) \\
&= \frac{0}{P(B)} && \text{(Since } P \text{ is a probability measure, } P(\emptyset) = 0) \\
&= 0 && \text{(Since } P(B) > 0 \text{ by assumption)}
\end{aligned}$$

This property holds.

2. Non-negativity: For any event $A \in \mathcal{A}$, we have:

$$\tilde{P}(A) = \frac{P(A \cap B)}{P(B)} \quad \text{(Definition of Conditional Probability)}$$

The numerator, $P(A \cap B)$, must be greater than or equal to 0 because P is a probability measure and $A \cap B$ is an event in \mathcal{A} . The denominator, $P(B)$, is strictly greater than 0 by the problem statement. The ratio of a non-negative number and a positive number is always non-negative. Thus, $\tilde{P}(A) \geq 0$ for all $A \in \mathcal{A}$.

3. σ -additivity: This is the most crucial property. Let $(A_i)_{i \in \mathbb{N}}$ be a sequence of pairwise disjoint sets in \mathcal{A} (i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$). We need to show that $\tilde{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \tilde{P}(A_i)$.

$$\begin{aligned}
\tilde{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) && \text{(Definition of } \tilde{P}) \\
&= \frac{P((\bigcup_{i=1}^{\infty} A_i) \cap B)}{P(B)} && \text{(Def. of Conditional Probability)} \\
&= \frac{P(\bigcup_{i=1}^{\infty} (A_i \cap B))}{P(B)} && \text{(Distributive law of sets)}
\end{aligned}$$

Now, consider the sets $(A_i \cap B)$. Since the A_i are pairwise disjoint, the sets $(A_i \cap B)$ are also pairwise disjoint. Because P is a probability measure, it is σ -additive. We can thus rewrite the numerator:

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

Substituting this back, we get:

$$\begin{aligned}
\tilde{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} && (\sigma\text{-additivity of } P) \\
&= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} && \text{(Pulling the constant denominator into the sum)} \\
&= \sum_{i=1}^{\infty} P(A_i|B) && \text{(Def. of Conditional Probability)} \\
&= \sum_{i=1}^{\infty} \tilde{P}(A_i) && \text{(Definition of } \tilde{P})
\end{aligned}$$

The σ -additivity property holds.

4. Normalization: Finally, we check if $\tilde{P}(\Omega) = 1$.

$$\begin{aligned}
\tilde{P}(\Omega) &= P(\Omega|B) && \text{(Definition of } \tilde{P}) \\
&= \frac{P(\Omega \cap B)}{P(B)} && \text{(Def. of Conditional Probability)} \\
&= \frac{P(B)}{P(B)} && \text{(Since } B \subseteq \Omega, \text{ we have } \Omega \cap B = B) \\
&= 1 && \text{(Since } P(B) > 0)
\end{aligned}$$

All four properties are satisfied, so \tilde{P} is a valid probability measure on (Ω, \mathcal{A}) . The space $(B, \mathcal{A}|_B, \tilde{P}|_B)$ is called the trace, where $\mathcal{A}|_B$ is the induced σ -algebra on B [5].

3 Part (ii): Law of Total Probability

Statement

Let $\{B_i \mid i \in I\} \subseteq \mathcal{A}$ be a countable, measurable partition [4] of $B \in \mathcal{A}$ (i.e., $\bigcup_{i \in I} B_i = B$ and $B_i \cap B_j = \emptyset$ for $i \neq j$). Show that for any $A \in \mathcal{A}$,

$$P(A \cap B) = \sum_{i \in I} P(A|B_i)P(B_i)$$

This is a slight generalization of **Theorem 1.65 (ii)** from the script.

Proof Walkthrough

We will start with the right-hand side (RHS) and show that it simplifies to the left-hand side (LHS). For this to be well-defined, we assume $P(B_i) > 0$ for all $i \in I$. If $P(B_i) = 0$ for some i , the term $P(A|B_i)P(B_i)$ is taken to be 0.

$$\begin{aligned}
\sum_{i \in I} P(A|B_i)P(B_i) &= \sum_{i \in I} \frac{P(A \cap B_i)}{P(B_i)} P(B_i) && \text{(Def. of Conditional Probability)} \\
&= \sum_{i \in I} P(A \cap B_i) && \text{(Canceling } P(B_i))
\end{aligned}$$

Now, because the sets B_i are pairwise disjoint, the sets $(A \cap B_i)$ must also be pairwise disjoint. Since P is a probability measure, it is σ -additive. Therefore, we can write the sum as the probability of the union:

$$\begin{aligned}
\sum_{i \in I} P(A \cap B_i) &= P\left(\bigcup_{i \in I} (A \cap B_i)\right) && (\sigma\text{-additivity of } P) \\
&= P\left(A \cap \left(\bigcup_{i \in I} B_i\right)\right) && \text{(Distributive law of sets)}
\end{aligned}$$

By the problem statement, $\{B_i\}$ is a partition of B , which means $\bigcup_{i \in I} B_i = B$. Substituting this in gives:

$$P\left(A \cap \left(\bigcup_{i \in I} B_i\right)\right) = P(A \cap B)$$

This is the LHS of the equation, completing the proof.

4 Part (iii): Bayes' Rule

Statement

For each $A \in \mathcal{A}$ with $P(A) > 0$ and every measurable partition $\{E_i \mid i \in I\}$ of Ω with $P(E_i) > 0$ for all $i \in I$, show that:

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_{j \in I} P(A|E_j)P(E_j)} \quad \text{for all } i \in I$$

This is exactly **Theorem 1.65 (iii)** from the script.

Proof Walkthrough

The proof consists of expanding the definition of $P(E_i|A)$ and then rewriting the numerator and denominator.

Step 1: Expand the definition We start with the definition of conditional probability:

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

Step 2: Rewrite the Numerator The joint probability $P(E_i \cap A)$ can be expressed using the definition of conditional probability in the other direction:

$$P(E_i \cap A) = P(A|E_i)P(E_i)$$

This is sometimes called the "chain rule" or "multiplication rule" for two events.

Step 3: Rewrite the Denominator To rewrite the denominator $P(A)$, we use the Law of Total Probability which we just proved in Part (ii). Since $\{E_j \mid j \in I\}$ is a partition of the entire sample space Ω , we can set $B = \Omega$ in our formula from part (ii):

$$P(A) = P(A \cap \Omega) = \sum_{j \in I} P(A|E_j)P(E_j)$$

Step 4: Combine the pieces Now we substitute our expanded forms for the numerator and denominator back into the original expression for $P(E_i|A)$:

$$P(E_i|A) = \frac{P(E_i \cap A)}{P(A)} = \frac{P(A|E_i)P(E_i)}{\sum_{j \in I} P(A|E_j)P(E_j)}$$

This is precisely Bayes' Rule. The proof is complete.

Further Explanations

- [1] **Probability Space** (Ω, \mathcal{A}, P) : This is the fundamental triplet of modern probability theory, as defined in **Def. 1.18**.
- Ω : The **sample space**, which is the set of all possible outcomes of a random experiment (e.g., {heads, tails} for a coin flip). It must be non-empty.
 - \mathcal{A} : A **σ -algebra** [3] on Ω , also called the **event space**. It is a collection of subsets of Ω that we consider to be "events" to which we can assign a probability. Not every subset of Ω is necessarily in \mathcal{A} , especially for uncountable sample spaces.
 - P : A **probability measure**, which is a function $P : \mathcal{A} \rightarrow [0, 1]$ that assigns a probability to each event in \mathcal{A} , satisfying the axioms we verified in Part (i).
- [2] **Conditional Probability**: Defined in **Def. 1.64**, the conditional probability of event A given that event B has occurred is:

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

This is only defined when $P(B) > 0$. Intuitively, it represents restricting our sample space to the outcomes in B and then renormalizing the probabilities so they sum to 1 over this new, smaller space.

- [3] **σ -Algebra**: As per **Def. 1.5**, a collection of subsets $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ is a σ -algebra if it satisfies three properties:

1. $\Omega \in \mathcal{A}$ (and thus $\emptyset \in \mathcal{A}$).
2. It is closed under complement: If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.
3. It is closed under countable unions: If $A_1, A_2, \dots \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

These properties ensure that if we can talk about certain events, we can also talk about their complements ("not A "), their countable unions ("at least one of the A_i "), and their countable intersections ("all of the A_i ").

- [4] **Measurable Partition**: A collection of sets $\{B_i \mid i \in I\}$ is a measurable partition of a set B if:

1. All sets are measurable: $B_i \in \mathcal{A}$ for all $i \in I$.
2. They are pairwise disjoint: $B_i \cap B_j = \emptyset$ for all $i \neq j$.
3. Their union is B : $\bigcup_{i \in I} B_i = B$.

This means the sets $\{B_i\}$ perfectly chop up the set B into non-overlapping pieces, and all those pieces are valid events.

- [5] **Induced σ -Algebra** $(\mathcal{A}|_B)$: From **Lemma 1.14**, given a measurable space (Ω, \mathcal{A}) and a non-empty subset $B \subseteq \Omega$, the induced σ -algebra on B is defined as:

$$\mathcal{A}|_B := \{A \cap B \mid A \in \mathcal{A}\}$$

This creates a valid σ -algebra on the new sample space B . It consists of all the ways that events from the original σ -algebra \mathcal{A} can "intersect with" or be "seen from within" the subset B .