

Session 1: Why Probability? Search and Belief

Movement I: Belief and Conditioning

Instructor: John Lipor

This is the first meeting of *Movement I: Belief and Conditioning*. A slogan will run through the whole course; we state it now and will repeat it at the start of every movement:

A probability model answers three questions: What is uncertain? What is observed? How should observation change belief and action?

Today we build the vocabulary for the first two questions and take our first step on the third. The session thesis: *probability is the calculus of belief, and conditioning is its most important operation*.

1 Launch

[**Thread A**] A vessel has gone down in one of four seafloor regions. We do not know which one; we have a sonar that sometimes fails to detect a wreck even when it passes directly over it. Where should we look, and (this is the subtle part) what should we conclude when a look comes back empty? This search problem is the spine of Movement I, and we open the course by playing it.

In-class search game (facilitation). Set four boxes at the front of the room, labeled regions 1–4. Announce the *prior* distribution for where the token (the “wreck”) is hidden: $(p_1, p_2, p_3, p_4) = (0.4, 0.3, 0.2, 0.1)$. Out of the class’s sight, roll to place the token according to this prior. The class now buys “looks.” A look at the *correct* box finds the token only if a rolled die shows 1–4; if the die shows 5 or 6 the look fails even though the token is there, so the *overlook probability* is $\alpha = 1/3$. A look at a *wrong* box never finds the token. The class votes on where to look; announce success or failure; repeat until found; tally the total number of looks.

Then play a *second* round, and after each failed look ask the room two questions: “*What do you believe now?*” and “*Did that failed look tell you anything?*” Steer the discussion until someone says out loud the point of the whole session: a failed look is not nothing: it is evidence, and it should move your belief away from the box you just searched and toward the others.

The game exposes three needs that organize the rest of today. We need a language for *what is uncertain* (a sample space and events). We need rules for combining and comparing the probabilities of events (the axioms and their consequences). And we need an operation that takes “the look in region 1 failed” and turns our prior into an updated belief. That operation is *conditioning*, and it is the heart of the session.

2 Development

2.1 The probability model

Every probabilistic statement is made relative to a model of “what could happen.”

Definition 1 (Sample space and events). *The **sample space** Ω is the set of all possible outcomes of an experiment. An **event** is a subset of Ω ; we say the event occurs when the realized outcome lies in it.*

In the search game, $\Omega = \{1, 2, 3, 4\}$ records which region hides the wreck, and “the wreck is in region 1 or 2” is the event $\{1, 2\}$. Because events are sets, the logical operations “and,” “or,” and “not” become the set operations $\cap$, $\cup$, and complement $A^c = \Omega \setminus A$.

The last ingredient is a way to measure how likely each event is. We ask only that this measurement obey three rules.

Definition 2 (Probability measure and axioms). *A **probability measure** $\mathbb{P}$ assigns to each event A a number $\mathbb{P}(A)$ such that*

1. (**nonnegativity**) $\mathbb{P}(A) \geq 0$ for every event A ;
2. (**normalization**) $\mathbb{P}(\Omega) = 1$;
3. (**countable additivity**) if $A_1, A_2, \dots$ are pairwise disjoint events ($A_i \cap A_j = \emptyset$ for $i \neq j$), then

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

Remark 1. *We state countable additivity because it is the standard axiom, but everything in this course concerns finitely many events at a time, and for that the finite version ($\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n \mathbb{P}(A_i)$ for disjoint A_i) is all we ever use. There is a genuine technical question hiding here: for uncountable Ω one cannot sensibly assign a probability to every subset, so “event” really means “member of a distinguished collection of measurable sets.” Pinning down that collection is the subject of measure theory. We will never need it, and this is the last we will say about it.*

2.2 Three consequences of the axioms

The axioms are deliberately spare; a great deal follows from them. We prove the three facts we will lean on all term.

Proposition 1. *For any events A, B :*

1. (complement) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;
2. (monotonicity) if $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$;
3. (inclusion–exclusion) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Proof. (1) A and A^c are disjoint and $A \cup A^c = \Omega$, so additivity and normalization give $\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\Omega) = 1$; rearrange.

(2) If $A \subseteq B$, then $B = A \cup (B \setminus A)$ is a disjoint union, so $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$, because $\mathbb{P}(B \setminus A) \geq 0$ by nonnegativity.

(3) Split both $A \cup B$ and B into disjoint pieces using A^c :

$$A \cup B = A \cup (B \cap A^c), \quad \text{so } \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c), \quad (2)$$

$$B = (B \cap A) \cup (B \cap A^c), \quad \text{so } \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \cap A^c). \quad (3)$$

Solving the second equation for $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ and substituting into the first gives the claim. $\square$

2.3 What does a probability mean?

Remark 2 (Two readings of $\mathbb{P}$). *There are two honest ways to read the number $\mathbb{P}(A)$. The frequency reading says $\mathbb{P}(A)$ is the long-run fraction of trials in which A occurs; it fits repeatable experiments like coin tosses. The degree-of-belief reading says $\mathbb{P}(A)$ measures how strongly a rational agent, given its information, expects A ; it fits one-off facts. The wreck is somewhere specific right now (there is nothing random about its location), yet we are ignorant of it, and $\mathbb{P}(\text{region } i)$ quantifies that ignorance. This course uses probability freely in the second sense, to reason about fixed but unknown quantities. The axioms and all their consequences hold verbatim under either reading, which is exactly why the same machinery serves both.*

2.4 Conditional probability

We reach the central operation. We know the prior probabilities of the four regions; a look then *observes* something (“region 1 came up empty”), and we want the probability of each region *in light of that observation*. The tool is conditional probability.

Definition 3 (Conditional probability). *For events A and B with $\mathbb{P}(B) > 0$, the **conditional probability of A given B** is*

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

Key idea. Conditioning is *restriction and renormalization*. Learning that B occurred discards every outcome outside B : the event B becomes the new sample space. We keep only the part of A that lives inside B , namely $A \cap B$, and we divide by $\mathbb{P}(B)$ so that the surviving world again has total probability 1. The division is the renormalization; without it the conditional “probabilities” would not sum to one.

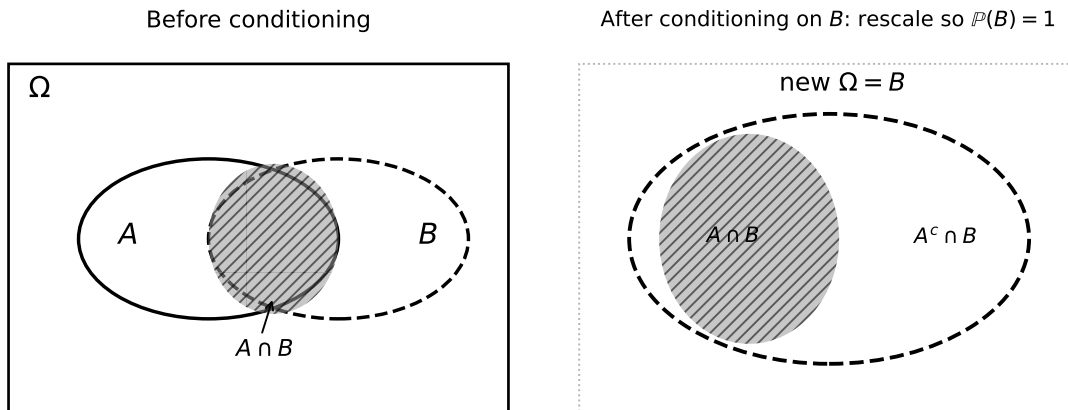


Figure 1: Conditioning on B restricts the sample space to B (the new Ω) and renormalizes: $\mathbb{P}(A|B)$ is the share of the rescaled B -world occupied by $A \cap B$.

2.5 The chain rule

Multiplying the definition through by $\mathbb{P}(B)$ rearranges it into a form we will use constantly to *build* the probability of a joint occurrence out of a probability and a conditional probability.

Proposition 2 (Chain rule). *For events with the relevant conditioning probabilities positive,*

$$\mathbb{P}(A \cap B) = \mathbb{P}(B) \mathbb{P}(A|B), \quad (5)$$

and for three events,

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \mathbb{P}(B|A) \mathbb{P}(C|A \cap B). \quad (6)$$

Proof. The two-event form is the definition of $\mathbb{P}(A|B)$ cleared of its denominator. For three events, apply the two-event form twice: writing $D = A \cap B$, $\mathbb{P}(D \cap C) = \mathbb{P}(D) \mathbb{P}(C|D)$, and then $\mathbb{P}(D) = \mathbb{P}(A) \mathbb{P}(B|A)$. $\square$

2.6 First contact with the search update

We now do, with nothing but the definition of conditional probability, the computation the class was approximating by intuition during the game. To keep the arithmetic in view we use just two regions.

Suppose the wreck is in region 1 or region 2, with prior $\mathbb{P}(\text{cell 1}) = 0.7$ and $\mathbb{P}(\text{cell 2}) = 0.3$. We look in cell 1. If the wreck is there, the look fails with the overlook probability $\alpha = 0.5$; if the wreck is *not* there, a look in cell 1 certainly fails to find it. Let F be the event “the look in cell 1 fails.” We want $\mathbb{P}(\text{cell 1} | F)$.

First assemble $\mathbb{P}(F)$ by splitting on the two cases (this is the finite additivity of Proposition 1 in action, applied to the disjoint pieces $F \cap \{\text{cell 1}\}$ and $F \cap \{\text{cell 2}\}$, each rewritten

with the chain rule):

$$\begin{aligned}
\mathbb{P}(F) &= \mathbb{P}(F \cap \{\text{cell 1}\}) + \mathbb{P}(F \cap \{\text{cell 2}\}) \\
&= \mathbb{P}(\text{cell 1}) \mathbb{P}(F | \text{cell 1}) + \mathbb{P}(\text{cell 2}) \mathbb{P}(F | \text{cell 2}) \\
&= (0.7)(0.5) + (0.3)(1) = 0.35 + 0.30 = 0.65.
\end{aligned} \tag{7}$$

Now apply the definition of conditional probability, using $\mathbb{P}(\text{cell 1} \cap F) = \mathbb{P}(\text{cell 1}) \mathbb{P}(F | \text{cell 1}) = 0.35$:

$$\mathbb{P}(\text{cell 1} | F) = \frac{\mathbb{P}(\text{cell 1} \cap F)}{\mathbb{P}(F)} = \frac{0.35}{0.65} \approx 0.538. \tag{8}$$

The belief in cell 1 has fallen from 0.70 to 0.538: *a failed look is evidence against the searched cell*, exactly as the game suggested. The mechanic is worth naming because we will repeat it all term: *restrict to the F-world, then renormalize*. We have not invoked any named theorem, only the definition. Session 2 gives this mechanic a name (Bayes' rule) and an industrial-strength formula for any number of cells.

3 Worked examples

3.1 (a) The two-cell failed-search update, line by line

We record the same computation as a template, justifying every line.

$$\mathbb{P}(F | \text{cell 1}) = \alpha = 0.5 \quad (\text{look of correct cell fails w.p. } \alpha) \tag{9}$$

$$\mathbb{P}(F | \text{cell 2}) = 1 \quad (\text{look of a wrong cell always fails}) \tag{10}$$

$$\mathbb{P}(F) = (0.7)(0.5) + (0.3)(1) = 0.65 \quad (\text{split on cases; chain rule}) \tag{11}$$

$$\mathbb{P}(\text{cell 1} | F) = \frac{(0.7)(0.5)}{0.65} = \frac{0.35}{0.65} \approx 0.538 \quad (\text{definition of conditioning}) \tag{12}$$

$$\mathbb{P}(\text{cell 2} | F) = \frac{(0.3)(1)}{0.65} = \frac{0.30}{0.65} \approx 0.462 \quad (\text{definition of conditioning}) \tag{13}$$

As a check the two posteriors sum to 1, as they must after renormalization. Note the qualitative pattern that Session 2 will prove in general: the searched cell's probability strictly *decreased* ($0.70 \rightarrow 0.538$) and the unsearched cell's probability *increased* ($0.30 \rightarrow 0.462$). Probability flowed out of the box we searched and into the box we did not.

3.2 (b) A conditional where $\mathbb{P}(A | B) \neq \mathbb{P}(B | A)$

Draw one card from a standard 52-card deck. Let $A = \{\text{the card is a King}\}$ and $B = \{\text{the card is a face card (Jack, Queen, or King)}\}$. Every King is a face card, so $A \cap B = A$, and thus $\mathbb{P}(A \cap B) = \mathbb{P}(A) = 4/52$. Then

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{4/52}{12/52} = \frac{4}{12} = \frac{1}{3}, \tag{14}$$

$$\mathbb{P}(B | A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{4/52}{4/52} = 1. \tag{15}$$

The two conditionals are wildly different: given a face card, the chance it is a King is $1/3$; given a King, the chance it is a face card is a certainty. The direction of conditioning matters. In general the two are related by $\mathbb{P}(A | B) = \mathbb{P}(B | A) \mathbb{P}(A) / \mathbb{P}(B)$ (rearrange the chain rule), so they agree only when $\mathbb{P}(A) = \mathbb{P}(B)$.

Pitfall. Do not confuse $\mathbb{P}(A | B)$ with $\mathbb{P}(B | A)$. In a courtroom, “the probability of this DNA match if the defendant is innocent is one in a million” ($\mathbb{P}(\text{match} | \text{innocent})$ small) is routinely misheard as “the probability the defendant is innocent given the match is one in a million” ($\mathbb{P}(\text{innocent} | \text{match})$), a different quantity that also depends on how many other people could have matched. Swapping the two is the *prosecutor’s fallacy*.

Pitfall. The definition of $\mathbb{P}(A | B)$ requires $\mathbb{P}(B) > 0$; conditioning on an event of probability zero is, for now, undefined. This looks like a mere technicality, but it becomes a real obstacle once we condition on continuous observations like a range reading of exactly 103.7 m, where every single value has probability zero. We flag the issue now and resolve it properly in Session 8.

4 Consolidation

Work these in class.

1. (*Mechanical.*) A wreck is in cell 1 with prior 0.6 or cell 2 with prior 0.4. A look in cell 1 has overlook probability $\alpha = 1/3$. Let F be “the look in cell 1 fails.” Compute $\mathbb{P}(F)$ and $\mathbb{P}(\text{cell 1} | F)$.
Answer: $\mathbb{P}(F) = (0.6)(1/3) + (0.4)(1) = 0.6$; $\mathbb{P}(\text{cell 1} | F) = 0.2/0.6 = 1/3$.
2. (*Conceptual.*) In Example (b), you found $\mathbb{P}(A | B) = 1/3$ while $\mathbb{P}(B | A) = 1$. Without any new computation, explain in one sentence why $\mathbb{P}(B | A)$ *had* to equal 1, referring to the sets A and B .
Answer: Because $A \subseteq B$ (every King is a face card), observing A guarantees B , so $\mathbb{P}(B | A) = 1$.
3. (*Conceptual.*) In the two-cell game, we found a failed look *lowers* the searched cell’s probability. Suppose instead the sensor were perfect ($\alpha = 0$). What is $\mathbb{P}(\text{cell 1} | F)$ then, and does the answer make sense?
Answer: $\mathbb{P}(F | \text{cell 1}) = 0$, so $\mathbb{P}(\text{cell 1} \cap F) = 0$ and $\mathbb{P}(\text{cell 1} | F) = 0$: a perfect sensor that reports failure proves the wreck is not in cell 1.

5 Synthesis

Return to the slogan. *What is uncertain?* The wreck’s region, modeled by a sample space and a prior over its events. *What is observed?* The outcome of a look, an event like F . *How should observation change belief?* By conditioning: restrict to the world in which the observation happened, then renormalize. That single operation converted a prior of 0.70 on cell 1 into a posterior of 0.538 after one failed look, and it did so using nothing beyond

the definition $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ and the axioms. Conditioning is the calculus of belief promised in the session thesis; every remaining session is, in one way or another, a new setting for this same move. Next time we give the restriction-and-renormalize mechanic its name and a formula that handles any number of cells and any sensor at once.

Self-check questions

1. Why must $\mathbb{P}(B) > 0$ in the definition of $\mathbb{P}(A|B)$?

Answer: The definition divides by $\mathbb{P}(B)$; conditioning on a probability-zero event is undefined (revisited in Session 8).

2. After a failed look in cell 1 (two-cell game, prior 0.7/0.3, $\alpha = 0.5$), did cell 2's probability rise or fall, and to what value?

Answer: It rose, from 0.30 to $0.30/0.65 \approx 0.462$.

3. Give one situation in which $\mathbb{P}(A|B) = \mathbb{P}(A)$, and say what it means.

Answer: When $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, i.e., A and B are independent: observing B leaves belief in A unchanged (the topic of Session 3).