

Session 2: Bayes' Rule and Sequential Updating

Movement I: Belief and Conditioning

Instructor: John Lipor

Session 1 ended with a promise. We had turned a prior of 0.70 on a cell into a posterior of 0.538 after one failed look, using nothing but the definition $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$, and we called the move *restriction and renormalization*. Today we give that move its name, *Bayes' rule*, and a formula that updates a belief over any number of cells, for any sensor, in one line. The session thesis: *Bayes' rule is the unique consistent way to move belief when evidence arrives, and applied repeatedly it is a learning algorithm.*

1 Launch

[**Thread A**] Reopen the state of the Session 1 search game. Suppose the class looked in cell 1 twice and came up empty both times. Watch what happened to the votes: on the first ballot the room still favored cell 1 (it had the largest prior), but after the second failure the votes drifted toward cell 2. Nobody in the room was doing arithmetic; they were running an approximation to a law. Our job today is to write that law down exactly, prove it, and then push it further than intuition can follow: what does belief do after ten failed looks, or a hundred? By the end of the hour you will be able to answer, on paper, questions the voting could only gesture at.

Everything rests on two results that upgrade last session's definition into a machine: the *law of total probability*, which assembles the probability of an observation from its possible causes, and *Bayes' rule*, which runs that assembly backward to update belief about the cause.

2 Development

2.1 The law of total probability

We first need vocabulary for “a complete list of mutually exclusive possibilities.”

Definition 1 (Partition). *A collection of events $\{B_1, \dots, B_n\}$ is a **partition** of Ω if the B_i are pairwise disjoint ($B_i \cap B_j = \emptyset$ for $i \neq j$) and together they exhaust the sample space ($\bigcup_{i=1}^n B_i = \Omega$).*

In the search problem the events $B_i = \{\text{the wreck is in cell } i\}$ form a partition: the wreck is in exactly one cell.

Theorem 1 (Law of total probability). *Let $\{B_1, \dots, B_n\}$ be a partition of Ω with $\mathbb{P}(B_i) > 0$ for every i . Then for any event A ,*

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A | B_i) \mathbb{P}(B_i). \quad (1)$$

Proof. Because the B_i partition Ω , the sets $A \cap B_i$ are pairwise disjoint and their union is $A = A \cap \Omega = A \cap \bigcup_i B_i = \bigcup_i (A \cap B_i)$. Finite additivity gives $\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \cap B_i)$, and the chain rule rewrites each term as $\mathbb{P}(A \cap B_i) = \mathbb{P}(A | B_i) \mathbb{P}(B_i)$. $\square$

Key idea. The law of total probability is a bookkeeping identity: to find how likely an observation A is overall, average its conditional probability $\mathbb{P}(A | B_i)$ over every possible state of the world, weighting each by how likely that state is. It is how we compute the denominator in what follows.

2.2 Bayes' rule

Total probability runs “causes $\rightarrow$ observation.” We want the reverse: an observation has arrived, and we want to revise belief about its cause.

Theorem 2 (Bayes' rule). *Let $\{B_1, \dots, B_n\}$ be a partition of Ω with each $\mathbb{P}(B_i) > 0$, and let A be an event with $\mathbb{P}(A) > 0$. Then for each i ,*

$$\mathbb{P}(B_i | A) = \frac{\mathbb{P}(A | B_i) \mathbb{P}(B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A | B_i) \mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A | B_j) \mathbb{P}(B_j)}. \quad (2)$$

Proof. The chain rule gives $\mathbb{P}(A \cap B_i)$ two ways: $\mathbb{P}(B_i | A) \mathbb{P}(A) = \mathbb{P}(A | B_i) \mathbb{P}(B_i)$. Divide by $\mathbb{P}(A) > 0$ for the first equality; the second substitutes the law of total probability (Theorem 1) for $\mathbb{P}(A)$. $\square$

The four pieces of (2) carry names we will use for the rest of the course. Read B_i as a hypothesis about the world and A as the data:

$$\underbrace{\mathbb{P}(B_i | A)}_{\text{posterior}} = \frac{\overbrace{\mathbb{P}(A | B_i)}^{\text{likelihood}} \cdot \overbrace{\mathbb{P}(B_i)}^{\text{prior}}}{\underbrace{\mathbb{P}(A)}_{\text{evidence}}}. \quad (3)$$

The **prior** $\mathbb{P}(B_i)$ is what we believed before the data; the **likelihood** $\mathbb{P}(A | B_i)$ is how well hypothesis i predicts the data we saw; the **evidence** $\mathbb{P}(A)$ is the total-probability normalizer; and the **posterior** $\mathbb{P}(B_i | A)$ is the updated belief. In three words: *posterior* $\propto$ *prior* $\times$ *likelihood*, with the evidence fixing the constant so the posterior sums to one.

Pitfall. The likelihood $\mathbb{P}(A | B_i)$ is a function of the hypothesis i with the data A held fixed. It is *not* a probability distribution over the hypotheses: the numbers $\mathbb{P}(A | B_1), \dots, \mathbb{P}(A | B_n)$ need not sum to 1, and there is no reason they should. (In the search game each is a single sensor probability like α or 1.) Only after multiplying by the prior and dividing by the evidence do we get something that sums to one across i .

2.3 The failed-search update

We now specialize Bayes' rule to the search problem and get a formula worth memorizing. Let the target lie in one of n cells with prior $(p_1, \dots, p_n)$. We search cell i . As in Session 1, a search of the cell that holds the target fails with the *overlook probability* α , while a search of a cell that does not hold the target always fails. Write $q = 1 - \alpha$ for the per-look detection probability, and let F_i be the event "the search of cell i fails."

Proposition 1 (Failed-search update). *With the setup above,*

$$\mathbb{P}(\text{cell } i \mid F_i) = \frac{\alpha p_i}{1 - q p_i}, \quad \mathbb{P}(\text{cell } j \mid F_i) = \frac{p_j}{1 - q p_i} \quad (j \neq i). \quad (4)$$

Proof. The cells partition the possibilities, so Bayes' rule applies with likelihoods $\mathbb{P}(F_i \mid \text{cell } i) = \alpha$ (the target is there but the look overlooks it) and $\mathbb{P}(F_i \mid \text{cell } j) = 1$ for $j \neq i$ (the look cannot find a target that is elsewhere). The evidence is, by the law of total probability,

$$\mathbb{P}(F_i) = \alpha p_i + \sum_{j \neq i} p_j = \alpha p_i + (1 - p_i) = 1 - (1 - \alpha) p_i = 1 - q p_i. \quad (5)$$

Dividing each likelihood-times-prior by this evidence gives (4). □

Corollary 1 (Probability flows out of the searched cell). *If $\alpha < 1$ and $p_i < 1$, then the searched cell's probability strictly decreases,*

$$\frac{\mathbb{P}(\text{cell } i \mid F_i)}{p_i} = \frac{\alpha}{1 - q p_i} < 1, \quad (6)$$

while every other cell's probability increases by the same factor $1/(1 - q p_i) > 1$. Belief leaves the cell we searched and is redistributed among the rest in proportion to their current probabilities.

Proof. The ratio for cell i is below 1 exactly when $\alpha < 1 - q p_i$, i.e. $q(1 - p_i) > 0$, which holds since $q = 1 - \alpha > 0$ and $p_i < 1$. For $j \neq i$ the ratio $\mathbb{P}(\text{cell } j \mid F_i)/p_j = 1/(1 - q p_i)$ exceeds 1 because $q p_i > 0$, and it does not depend on j . □

Pitfall. After writing down likelihood times prior, you must divide by the evidence. Skipping the division leaves you with the *unnormalized* numbers αp_i and p_j , which do not sum to one and are not probabilities. The evidence $1 - q p_i$ is exactly the sum of those numerators; dividing by it is the renormalization step from Session 1.

2.4 Repeated failed searches

What if we search the same cell again and again and keep failing? Model the looks as *conditionally independent given the target's location*: if the target is in cell i , each look fails independently with probability α , so m looks all fail with probability α^m ; if the target is elsewhere, all m looks fail with probability 1. (This conditional-independence assumption is the quiet workhorse of the whole course; Session 3 examines it directly.)

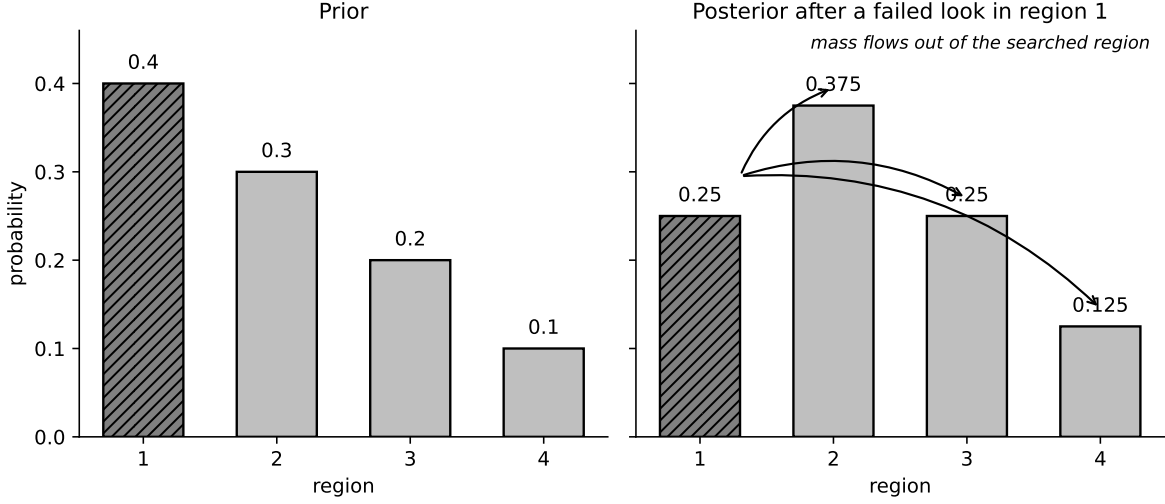


Figure 1: The failed-search update of Proposition 1 with prior $(0.4, 0.3, 0.2, 0.1)$ and $\alpha = 0.5$: one empty look in region 1 moves mass out of region 1 and into the others by the common factor $1/(1 - q p_1) = 1.25$.

Proposition 2 (Vanishing belief under repeated failure). *After m failed searches of cell i , starting from prior p_i ,*

$$\mathbb{P}(\text{cell } i \mid m \text{ failures}) = \frac{p_i \alpha^m}{p_i \alpha^m + (1 - p_i)}. \quad (7)$$

For any $\alpha < 1$ this tends to 0 as $m \rightarrow \infty$.

Proof. Group the world into the two hypotheses “cell i ” (prior p_i) and “not cell i ” (prior $1 - p_i$). The likelihood of m failures is α^m under the first and 1 under the second, so Bayes’ rule gives (7). As $m \rightarrow \infty$, $\alpha^m \rightarrow 0$, so the numerator $\rightarrow 0$ while the denominator $\rightarrow 1 - p_i > 0$; hence the ratio $\rightarrow 0$. $\square$

One reading, worth stating aloud: unless the sensor is perfect ($\alpha = 0$) the belief drops fast, and unless the target is certainly in cell i to begin with, *persisting in searching a cell that keeps coming up empty is eventually irrational*. The only way repeated failure fails to erode belief is $\alpha = 1$, a sensor that never detects anything and therefore whose failures carry no information.

2.5 The odds form

For two hypotheses, Bayes’ rule has an especially clean shape. Define the **prior odds** of hypothesis B against its complement as $\mathbb{P}(B)/\mathbb{P}(B^c)$, and likewise the **posterior odds** $\mathbb{P}(B \mid A)/\mathbb{P}(B^c \mid A)$. Dividing the two instances of Bayes’ rule (2) (for B and for B^c) cancels the shared evidence $\mathbb{P}(A)$:

$$\underbrace{\frac{\mathbb{P}(B \mid A)}{\mathbb{P}(B^c \mid A)}}_{\text{posterior odds}} = \underbrace{\frac{\mathbb{P}(B)}{\mathbb{P}(B^c)}}_{\text{prior odds}} \cdot \underbrace{\frac{\mathbb{P}(A \mid B)}{\mathbb{P}(A \mid B^c)}}_{\text{likelihood ratio } L(A)}. \quad (8)$$

The update is now a single multiplication: *posterior odds* = *prior odds* $\times$ *likelihood ratio*. No normalizer appears, because odds are ratios and the evidence cancels.

The likelihood ratio is the evidence's whole content. Equation (8) says an observation influences belief *only* through its likelihood ratio $L(A) = \mathbb{P}(A|B)/\mathbb{P}(A|B^c)$. Two different observations with the same likelihood ratio move belief identically. This one number is the entire evidentiary payload of the data; it returns in Session 10 as the *detection statistic* that decides target-or-no-target.

3 Worked examples

3.1 (a) The four-region wreck search

Take the Session 1 prior $(p_1, p_2, p_3, p_4) = (0.4, 0.3, 0.2, 0.1)$ and $\alpha = 0.5$, so $q = 0.5$. We search region 1 and fail.

First failed look. The evidence is $1 - qp_1 = 1 - (0.5)(0.4) = 0.8$. Proposition 1 gives

$$\mathbb{P}(\text{cell 1} | F_1) = \frac{(0.5)(0.4)}{0.8} = \frac{0.20}{0.8} = 0.25, \quad \mathbb{P}(\text{cell 2} | F_1) = \frac{0.3}{0.8} = 0.375, \quad (9)$$

$$\mathbb{P}(\text{cell 3} | F_1) = \frac{0.2}{0.8} = 0.25, \quad \mathbb{P}(\text{cell 4} | F_1) = \frac{0.1}{0.8} = 0.125. \quad (10)$$

The four posteriors sum to 1, and every unsearched cell grew by the factor $1/0.8 = 1.25$ promised by Corollary 1. Notice already that region 2, at 0.375, has *overtaken* region 1, at 0.25: a single empty look was enough to unseat the favorite.

Second failed look in region 1. Treat the posterior above as the new prior; now $p_1 = 0.25$ and the evidence is $1 - (0.5)(0.25) = 0.875$. Then

$$\mathbb{P}(\text{cell 1} | F_1 F_1) = \frac{(0.5)(0.25)}{0.875} \approx 0.143, \quad \mathbb{P}(\text{cell 2} | F_1 F_1) = \frac{0.375}{0.875} \approx 0.429, \quad (11)$$

with regions 3 and 4 at ≈ 0.286 and ≈ 0.143 . Equivalently, the two-failure formula (7) gives region 1 directly as $\frac{0.4(0.5)^2}{0.4(0.5)^2 + 0.6} = \frac{0.1}{0.7} \approx 0.143$, a useful check.

When does region 2 overtake region 1? Region 2 passes region 1 after the *first* failed look (posteriors 0.375 vs. 0.25), and the gap only widens with more failures. The crossover happens as soon as the searched favorite has enough probability drained to fall below the runner-up, which here is immediate.

Historical sidebar. This exact update guided real searches. When the submarine USS *Scorpion* was lost in 1968, John Craven's team placed a prior over the seafloor and updated it as sonar sweeps came back empty, recovering the wreck near the posterior's peak. The 2011 search for Air France 447's flight recorders used the same Bayesian bookkeeping after earlier sweeps had failed. In both, a failed look was treated not as wasted effort but as data.

3.2 (b) A base-rate problem in odds form

A cognitive-radio node must decide whether a licensed channel is occupied. Occupancy is rare: the prior is $\mathbb{P}(\text{occupied}) = 0.02$. The node's detector has probability of detection $\text{PD} = \mathbb{P}(\text{flag} | \text{occupied}) = 0.95$ and probability of false alarm $\text{PFA} = \mathbb{P}(\text{flag} | \text{free}) = 0.05$. The detector raises a flag. How strongly should the node now believe the channel is occupied?

Work in odds. The prior odds of occupied against free are $0.02/0.98 = 1/49$, i.e. 1:49. The likelihood ratio of a flag is

$$L(\text{flag}) = \frac{\mathbb{P}(\text{flag} | \text{occupied})}{\mathbb{P}(\text{flag} | \text{free})} = \frac{0.95}{0.05} = 19. \quad (12)$$

By the odds form (8), the posterior odds are $\frac{1}{49} \times 19 = \frac{19}{49}$, i.e. 19:49, so

$$\mathbb{P}(\text{occupied} | \text{flag}) = \frac{19}{19 + 49} = \frac{19}{68} \approx 0.28. \quad (13)$$

Pitfall. A detector that is right 95% of the time fires, yet the channel is probably *still free*. That feels wrong until the odds form makes it obvious: the flag multiplied the odds by 19, but it started 49 times more likely to be free than occupied, and 19 cannot overcome 49 in a single step. Ignoring the small prior and reading $\text{PD} = 0.95$ as “95% sure” is *base-rate neglect*, the most common error in applied inference. The likelihood ratio is strong; the prior is stronger.

4 Consolidation

Work these in class.

1. (*Mechanical.*) Prior (0.5, 0.3, 0.2) over three cells, overlook probability $\alpha = 0.4$ (so $q = 0.6$). Search cell 1 and fail. Compute the full posterior.
Answer: evidence $1 - qp_1 = 1 - (0.6)(0.5) = 0.7$;
 posterior $\left(\frac{0.4 \cdot 0.5}{0.7}, \frac{0.3}{0.7}, \frac{0.2}{0.7}\right) = (0.286, 0.429, 0.286)$.
2. (*Odds form.*) A test for a rare fault has prior $\mathbb{P}(\text{fault}) = 0.01$, detection probability 0.9, false-alarm probability 0.1. The test fires. Give the posterior odds and probability of a fault.
Answer: prior odds 1:99, $L = 9$, posterior odds $9:99 = 1:11$, so $\mathbb{P}(\text{fault} | \text{fire}) = 9/108 = 1/12 \approx 0.083$.
3. (*Conceptual.*) In Proposition 2, what happens to the posterior of cell i under repeated failure if $\alpha = 1$? Explain in one sentence why this is the only exception to “belief erodes.”
Answer: it stays at p_i forever, because an $\alpha = 1$ sensor fails with probability 1 whether or not the target is present, so its failures have likelihood ratio 1 and carry no information.

5 Synthesis

Bayes' rule is the answer to the third question of the course slogan: *how should observation change belief?* It does so in one identity, posterior $\propto$ prior $\times$ likelihood, with the evidence renormalizing. Applied once, it turns a prior over cells into a posterior; applied repeatedly, with each posterior serving as the next step's prior, it is a learning algorithm that drove the belief in a fruitlessly searched cell to zero and, in the odds form, reduced an observation's whole effect to a single multiplier, the likelihood ratio. Every later session is a new home for this machinery: continuous data in Movement II, vectors in Movement IV, whole functions in Movement V, but always posterior $\propto$ prior $\times$ likelihood. One assumption, though, has been running silently underneath the repeated-search formula: that successive looks are independent *given* the target's location. Session 3 asks when that is legitimate, and what a belief is finally *for* once we have it, which is to choose an action.

Self-check questions

1. In Bayes' rule, which factor is a function of the hypothesis with the data fixed, and why need it not sum to one over hypotheses?

Answer: the likelihood $\mathbb{P}(A | B_i)$; it is a list of predictions of one fixed observation, not a distribution over i , so nothing forces $\sum_i \mathbb{P}(A | B_i) = 1$.

2. After a failed search of cell i , by what common factor do all the other cells' probabilities grow, and why is it the same for each?

Answer: by $1/(1 - qp_i)$; it is common because each unsearched cell's likelihood of the failure is 1, so only the shared evidence rescales them.

3. Prior odds 1:49 and a likelihood ratio of 19 give what posterior probability, and what is the lesson?

Answer: posterior odds 19:49, probability $19/68 \approx 0.28$; a strong likelihood ratio can still lose to a strong prior (base-rate neglect).