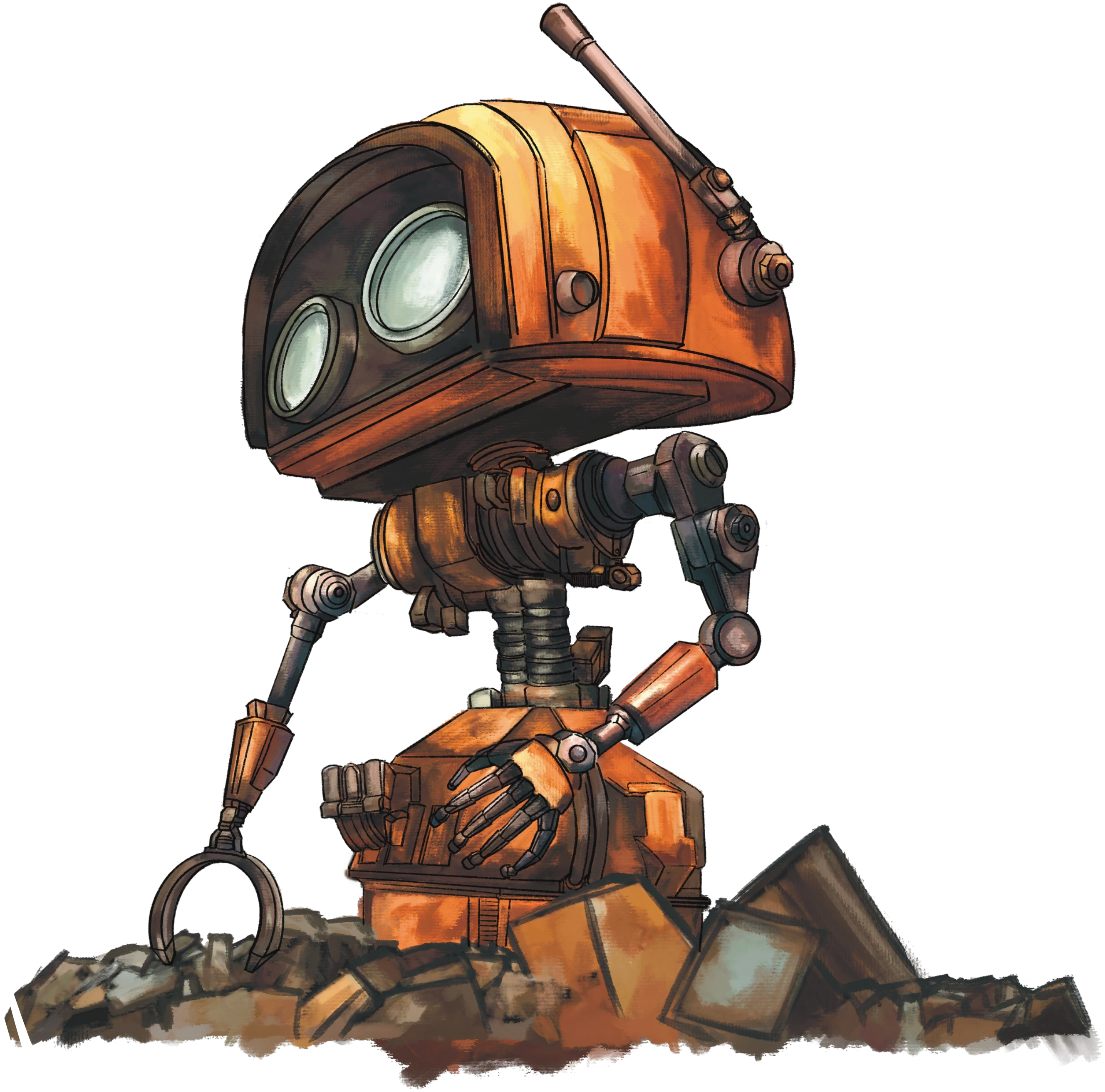
CS 3630, Fall 2026

Lecture 3: Introduction to Bayesian Thinking: from prior to posterior

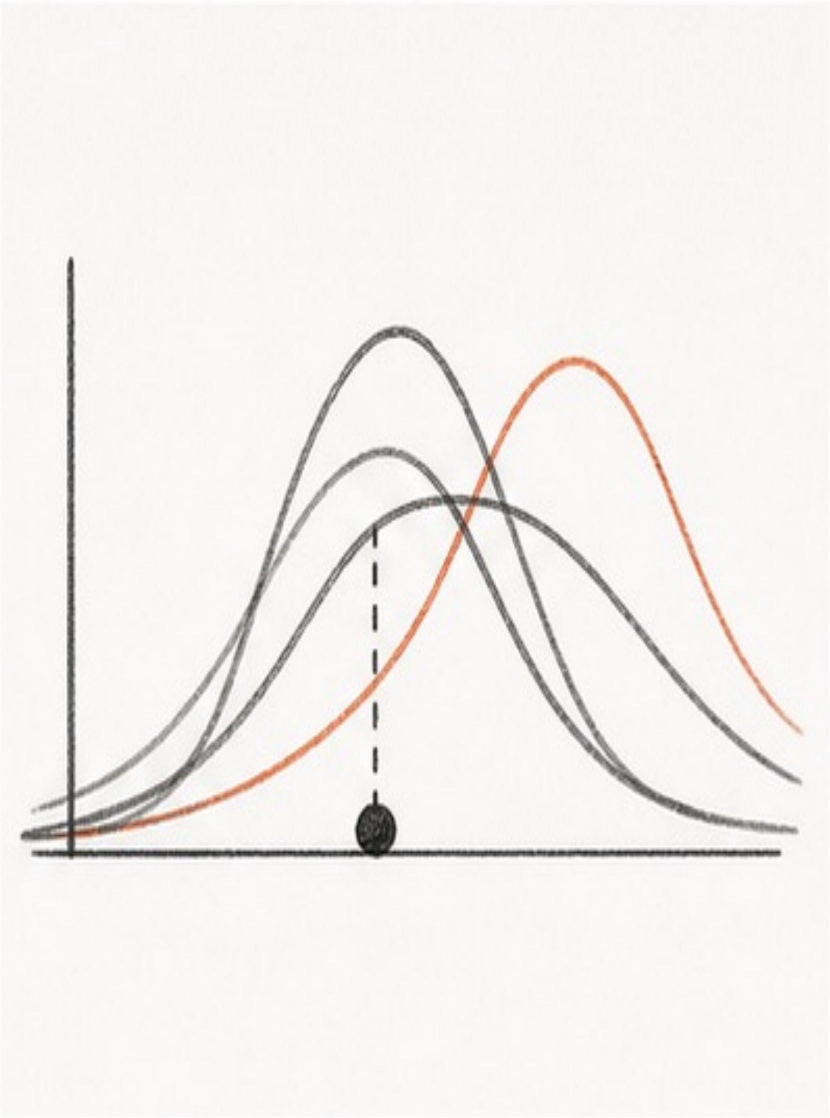


Why uncertainty needs a language

Robots, people, and software rarely know the world exactly. Probability lets us state what we believe, compare alternatives, and revise those beliefs when new information arrives.

Uncertainty is part of the model

- The world can be definite even when our information is incomplete.
- We represent several possible explanations at once.
- Probability says how strongly current information supports each explanation.
- New evidence should revise those probabilities



Belief can be a distribution

Suppose tomorrow can be sunny, cloudy, rainy, stormy, or snowy. We do not choose one outcome yet; we assign belief across all five.

Sample space: the possible outcomes.
Event: any subset of those outcomes.

A distribution assigns probability mass to outcomes.

Probability obeys three rules

1. Nonnegative: $P(A) \geq 0$.

2. Normalized: $P(\Omega) = 1$.

3. Additive: if A and B cannot both occur, $P(A \cup B) = P(A) + P(B)$.

*These rules make probability a consistent accounting system for **belief**.*

Events collect probability mass

Suppose tomorrow's weather has this distribution:

Outcome	Prob.
Sunny	0.55
Cloudy	0.25
Rain	0.15
Storm	0.04
Snow	0.01

Define W = {rain, storm, snow}.

This is the event "wet weather."

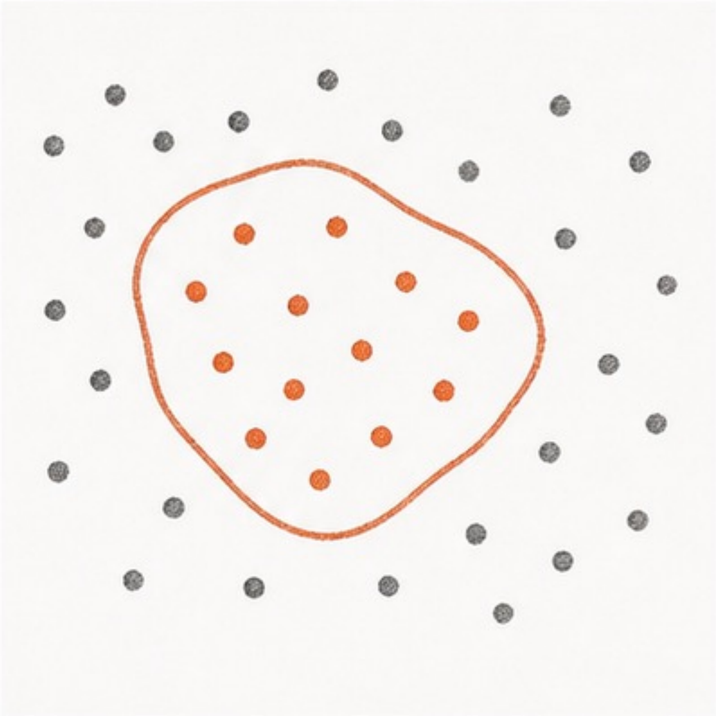
Question:

- What is P(W)?

Answer:

P(W) = 0.15 + 0.04 + 0.01 = 0.20

Events add the mass of their outcomes.



Priors: What did we believe before seeing data?

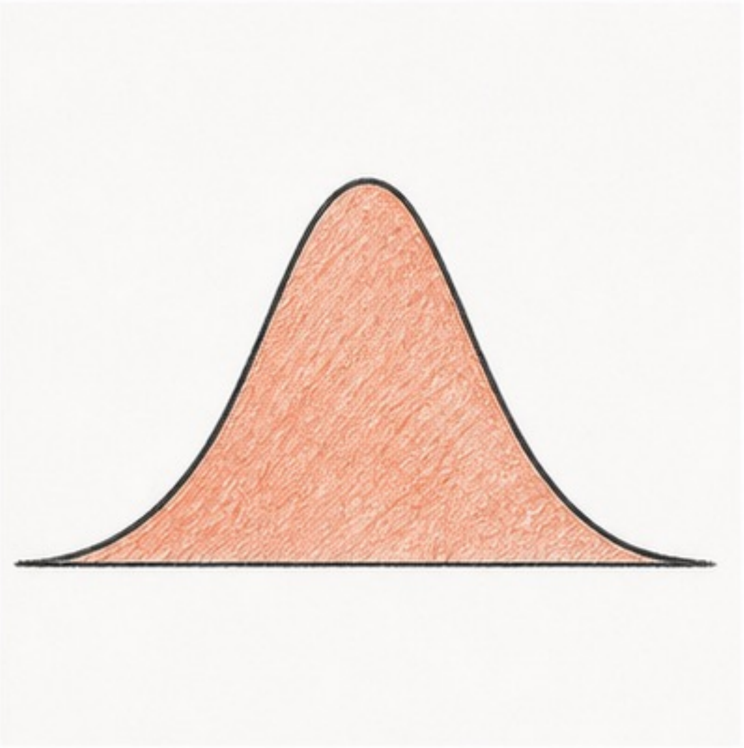
Bayesian reasoning starts ***before*** a new observation. It asks us to make the starting assumptions visible, encode them as a prior distribution, and then let evidence move them.

The prior is our starting belief

A prior probability distribution represents what is plausible before the current evidence is observed.

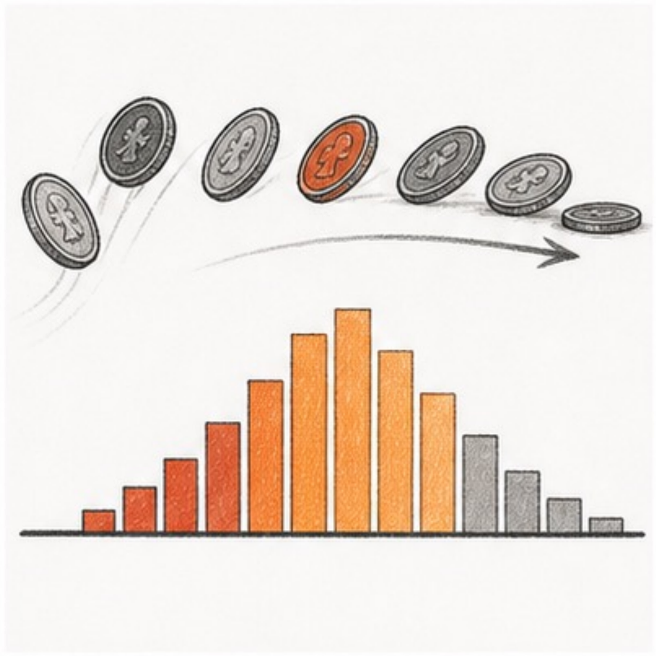
Priors may come from:

- previous observations
- domain knowledge
- physical constraints
- a deliberately neutral assumption



A prior is not permanent. It is a belief we agree to update.

Data can shape a prior



If an event appears m times in n comparable trials, m/n is a natural estimate of its probability. With more representative data, this empirical prior usually becomes more stable.

But the data source still matters: yesterday, another city, or another population may not describe today.

Start by stating what you believe

Bayesian thinking makes a simple discipline explicit:

1. List the competing hypotheses.
2. Assign a prior probability to each.
3. Describe what evidence each hypothesis predicts.
4. Update the probabilities after observing evidence.

The goal is to expose assumptions, then and **revise** *them.*

Evidence changes the odds

Evidence is useful only when different hypotheses predict it differently.
To update belief, we need a model of what we would expect to observe under each possible explanation.

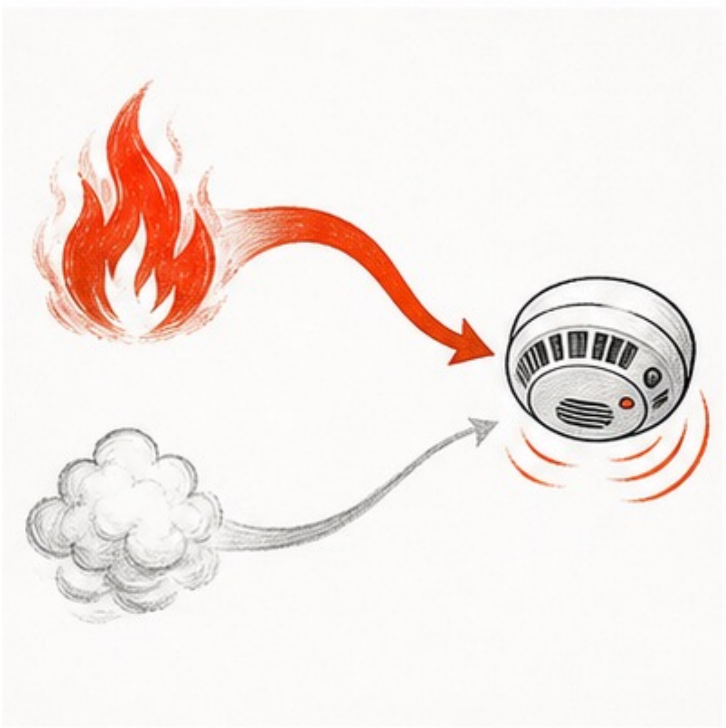
Evidence depends on the hypothesis

Two hypotheses:

F: there is a fire

¬F: there is no fire

Observation A: the alarm sounds



An alarm is evidence for F only because it is more likely during a fire than otherwise.

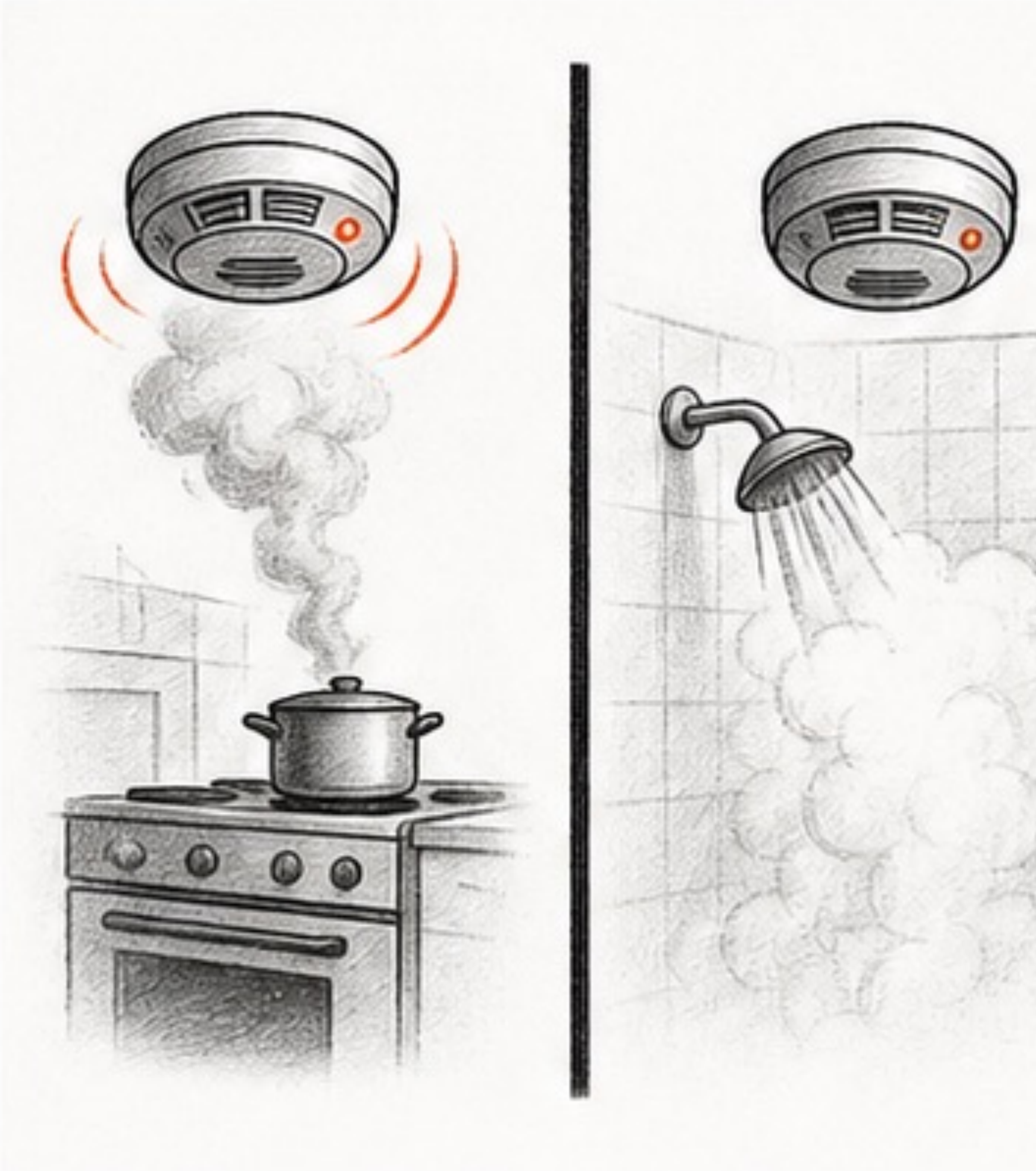
Conditional probability describes the strength of that dependence.

Conditional probability means "given"

P(A|F) means: the probability that the alarm sounds, given that there is a fire.

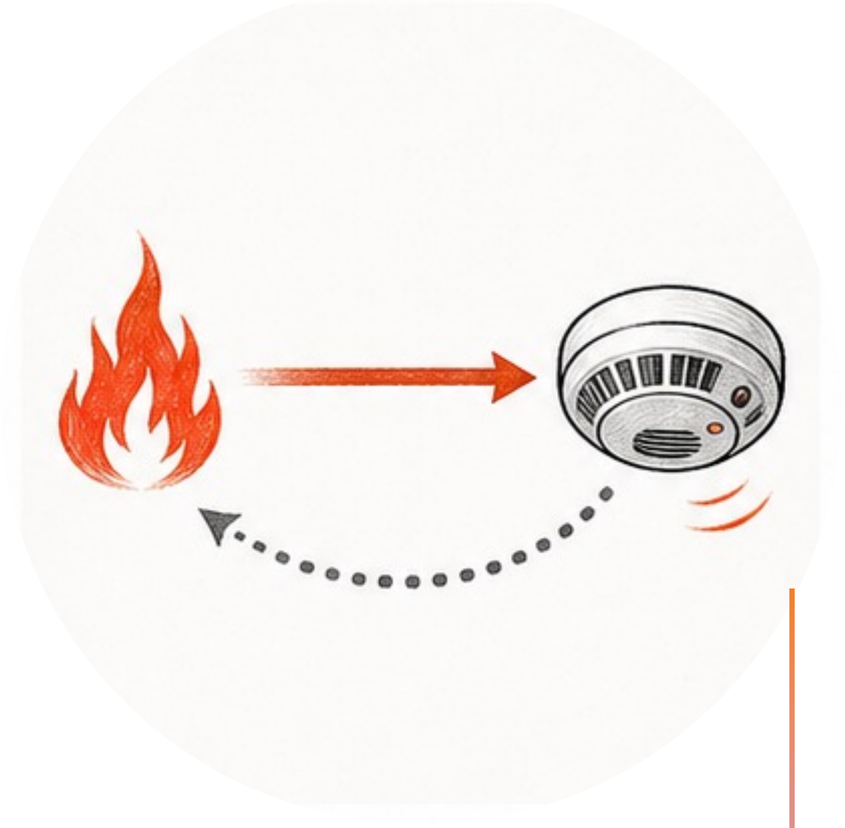
The condition after the vertical bar defines the world we are imagining.

Conditional probability connects hidden explanations to visible evidence.



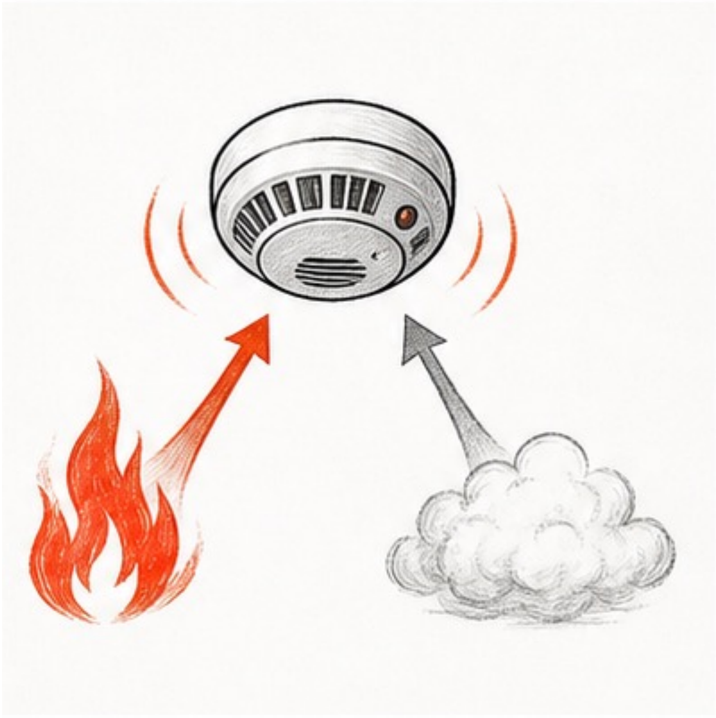
Do not reverse the condition

- ***P(alarm | fire) is not P(fire | alarm).***
- *The first asks how a sensor behaves when the cause is known.*
- *The second asks which cause is plausible after the observation.*
- ***Bayes' law connects the two questions.***



A noisy alarm asks an inverse question

Suppose the alarm has sounded. How likely is a fire now?



We know how the alarm behaves:

- P(A | F) = 0.90
- P(A | ¬F) = 0.05

We want the inverse: P(F | A)

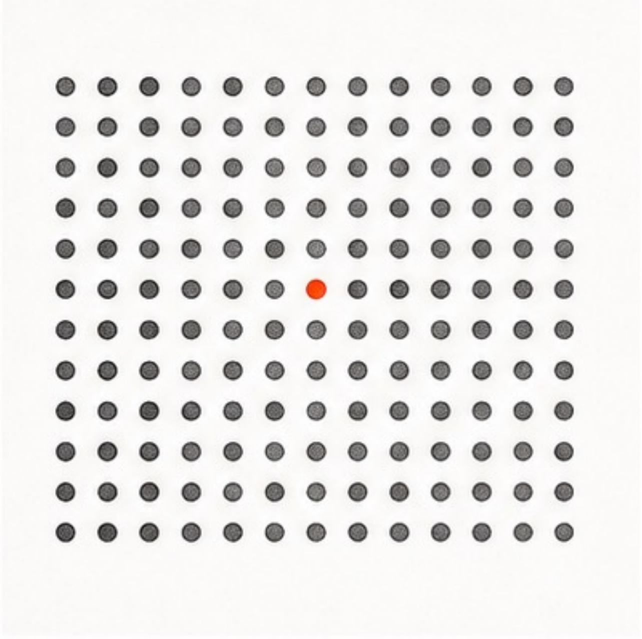
The answer depends on both the alarm and the base rate of fires.

Start with the base rate

Assume a fire is present only 1% of the times when this alarm could be observed:

State	Prior
Fire, F	0.01
No fire, ¬F	0.99
Total	1.00
Observation	A
Question	P(F | A)?

A 90%-sensitive alarm does not erase that fact.



The prior says fires are rare before hearing the alarm.

Base rates matter when one explanation is rare.
Next: ask how compatible the alarm is with each state.

Add the alarm model

For each possible state, ask how likely the observed alarm would be:

State	P(A | state)
Fire, F	0.90
No fire, ¬F	0.05
True positive	90%
False alarm	5%
Observed	A

Compute a weight for each state

- prior × likelihood

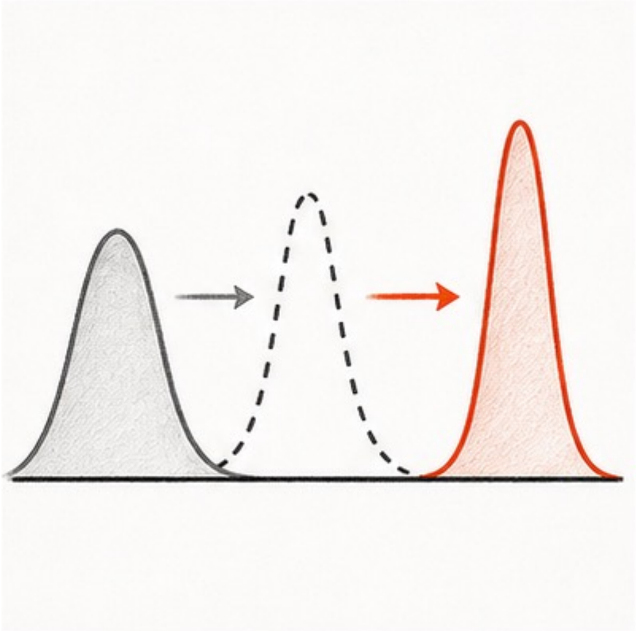
Fire: 0.01 × 0.90 = 0.0090
No fire: 0.99 × 0.05 = 0.0495

These are likelihoods: how well each hypothesis predicts the evidence.

Bayes’ law turns the model around

We know the prior $P(F)$ and the likelihood $P(A \mid F)$. Bayes' law gives the posterior probability of fire after the alarm.

$$P(F \mid A) = P(A \mid F)\, P(F) \,/\, P(A)$$



$P(A) = P(A \mid F)P(F) + P(A \mid \neg F)P(\neg F)$

The denominator adds all ways the evidence could have occurred, so the posterior sums to one.

Bayes' law has four ingredients

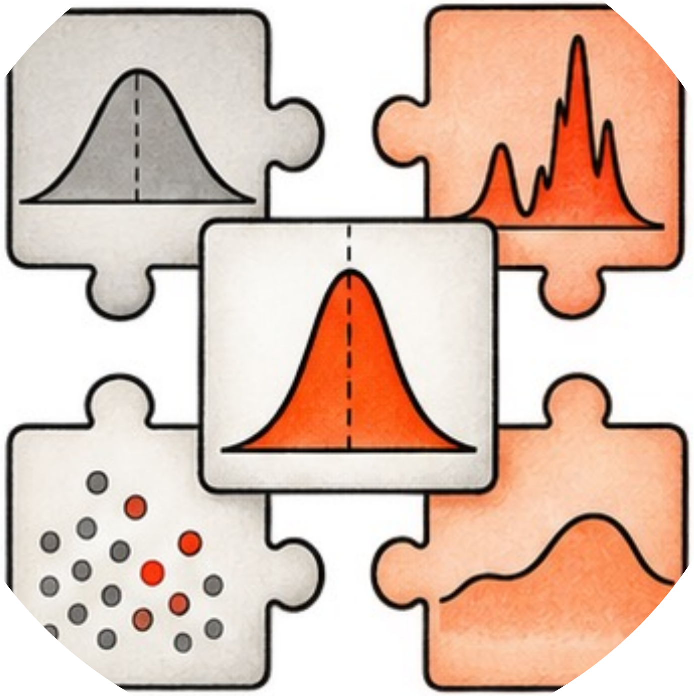
Prior — P(F): belief before the observation.

Likelihood — P(A | F): how strongly the hypothesis predicts the observation.

Evidence — P(A): probability of the observation overall.

Posterior — P(F | A): upgraded belief after the observation.

posterior = likelihood × prior ÷ evidence

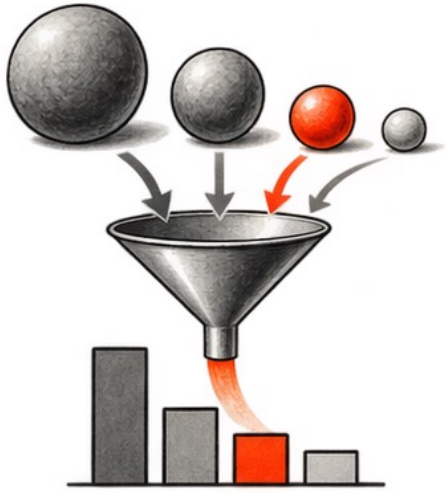


Compute weights, then normalize

Multiplying prior by likelihood gives unnormalized support. Dividing by the evidence converts those weights into a posterior distribution.

State	Prior	Like.	Weight	Post.	Read
F	0.010	0.90	0.0090	0.154	15.4%
¬F	0.990	0.05	0.0495	0.846	84.6%
P(A)	—	—	0.0585	—	evidence
Check	1.000	—	0.0585	1.000	sums
Now	1%	×	÷	15.4%	belief rises

Result	
0.0090	$0.01 \times 0.90 = 0.0090$
÷	$0.99 \times 0.05 = 0.0495$
0.0585	$P(A) = 0.0090 + 0.0495 = 0.0585$
0.1538	***$P(F \mid A) = 0.0090 / 0.0585 \approx 0.154$***



The base rate still matters

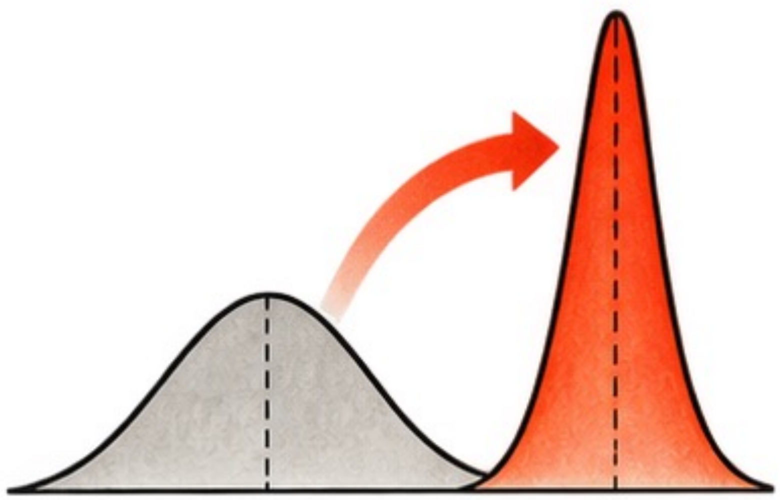
The alarm is strong evidence: belief in fire moves from 1% to about 15.4%.

But 90% sensitivity does not imply a 90% posterior. False alarms applied to the larger no-fire population still account for most alarms.

Bayes' law balances evidence against how common each explanation was beforehand.

Prior → posterior

Bayes' law is an upgrade rule. It combines what we believed before with how well each hypothesis predicted the evidence, producing a new distribution that reflects both.



The posterior is an upgraded belief

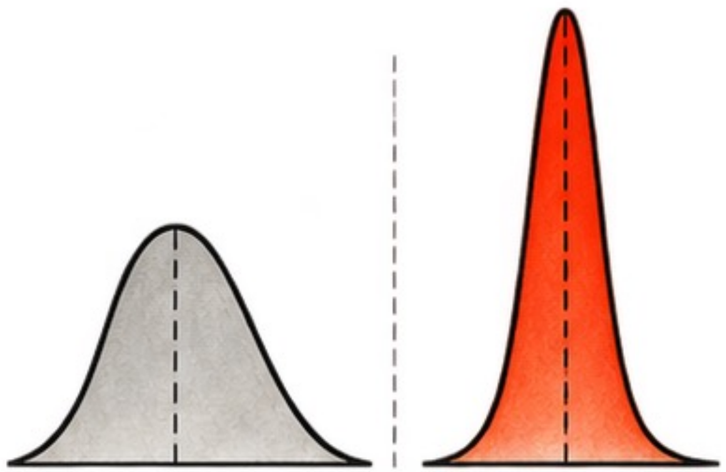
The posterior is not a new kind of object. It is another probability distribution—this time conditioned on the evidence we observed.

Before alarm:

- $P(F) = 0.010$
- $P(\neg F) = 0.990$

After alarm:

- $P(F \mid A) = 0.154$
- $P(\neg F \mid A) = 0.846$



Same hypotheses. New information. Revised probabilities.

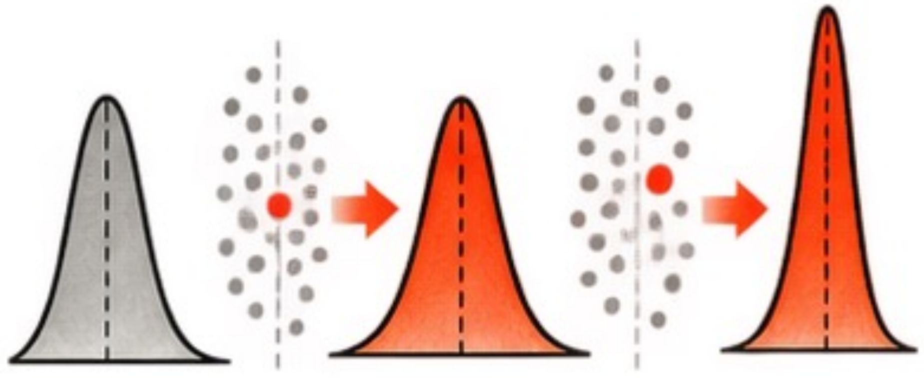
A posterior can become a new prior

Bayesian updating can repeat:

prior → observe evidence → posterior

That posterior becomes the prior for the next observation. Belief evolves as evidence accumulates.

Order does not matter when the evidence model is correct and observations are handled consistently.



Bayesian thinking allows for revision!

It replaces “I know” with a distribution—and replaces “I changed my mind” with an explicit update.

Hypotheses → prior → evidence
model → Bayes → posterior

Prior × likelihood ÷ evidence = posterior

Update again when new evidence arrives.

What comes next

- Build probability distributions over possible states.
- Model noisy measurements with conditional probability.
- Use Bayes' law to infer hidden state from visible evidence.
- Compare decisions using expected consequences.
- Learn model parameters from data.

Keep the update pattern in mind: prior + evidence → posterior.