Prior vs Posterior Explained

Teach me Prior vs Posterior like I am a school kid

Sure! Let's imagine you're a school kid who loves guessing games.

🎒 Imagine This:

You have a **mystery box**. Inside the box, there's **a candy**. It could be **red** \sum or **green** $\stackrel{\longleftarrow}{=}$.

Before opening it, you guess what color it might be.

That **first guess**, before you see anything, is called your *Prior belief*.

Step 1: Prior (What you think before you see anything)

You remember from the past:

"Hmm... usually this box has 7 red candies and 3 green candies."

So you think:

"I think there's a **higher chance** the candy is red!"

That's your **prior** — your belief before seeing any new clue.

Step 2: New Clue (Evidence)

Your friend peeks a little and says:

"I saw a little bit of the wrapper — it looked kind of red."

That's called evidence or data.

Step 3: Posterior (What you think after the clue)

Now you mix your **prior belief** (mostly red candies before) with the **new clue** (your friend saw red) and say:

"Now I'm even more sure the candy is red!"

That updated belief is your **OP** Posterior belief.

In Short:

Term	Simple Meaning	
Prior	What you believe before seeing clues.	
Evidence	What you see or learn.	
Posterior	What you believe <i>after</i> seeing the clues.	

Example:

- **Prior**: You think there's a 70% chance it's red.
- Evidence: Friend saw a red wrapper.
- **Posterior**: Now you think there's a 90% chance it's red.

That's how your belief gets smarter with new clues!

Want me to show this with pictures or math next?

Yes please

Awesome! Let's do both a picture and then some simple math, just like a school workbook!

Picture Explanation (Candy Box Story)

You used the **past knowledge** (prior) and the **new clue** (evidence) to **update your belief** (posterior).

🔢 Simple Math Example

Let's say:

- In the past, 70% of the candies were red
- You now get a clue: "I saw a red wrapper" (which is correct 90% of the time)

Let's calculate how sure you are **now** that the candy is red — the **posterior**.

Step 1: Prior

- P(Red) = 0.7
- P(Green) = 0.3

Step 2: Likelihood (how likely friend sees red if candy is really red/green)

- P(See red | Red candy) = 0.9
- P(See red | Green candy) = 0.2 (maybe red light confused your friend)

Step 3: Use Bayes' Rule (simplified)

Posterior =
$$\frac{\text{Prior} \times \text{Likelihood}}{\text{Total Chance of Seeing Red}}$$

Let's calculate the **numerator** for red:

$$P(\text{Red and saw red}) = 0.7 \times 0.9 = 0.63$$

And for green:

$$P(Green and saw red) = 0.3 \times 0.2 = 0.06$$

Total chance of seeing red:

$$P(\text{Saw red}) = 0.63 + 0.06 = 0.69$$

Now, final posterior:

$$P(\text{Candy is red} \mid \text{Saw red}) = \frac{0.63}{0.69} \approx 0.913$$

Example 2 Final Answer:

Even though you started with a **70% belief**, after the clue, you now believe there's a **91.3% chance** the candy is red!

Would you like a fun quiz or another example (like guessing if it rains)? 🌧 🍬

So does the bayes rule say,
Posterior * Chance of Seeing Red = Prior * Likelihood?
What exactly does the Bayes rule depict?

Great question

So does the bayes rule say,
Posterior * Chance of Seeing Red = Prior * Likelihood?
What exactly does the Bayes rule depict?

You're **very close** — and asking a super important question! **Solution**Let's clear it up with simple words and symbols.

What Bayes' Rule actually says:

Or in words:

Updated Belief =

(What you believed before) × (How well the clue fits) ÷ (How likely this clue is overall)

Rearranged version (like you wrote):

Yes — if you **multiply both sides** by the denominator, you get:

Posterior × Evidence = Prior × Likelihood

So yes, your version is algebraically valid!

But in most explanations, **Bayes' Rule is written like a ratio**, because it **tells how to update a belief**.

So What Does Bayes' Rule Depict?



"Given some clue or evidence, how should I change what I believe?"

It blends:

- What you knew before (Prior)
- What you *just saw* (**Likelihood**)
- And balances it with how surprising that clue is overall (Evidence)

@ Everyday Analogy

If I *already thought* it was likely to rain (prior), and now I *see black clouds* (evidence that's likely when rain is true), then my **belief in rain today should go up** (posterior)!

Bonus Insight:

This is the **core idea behind Bayesian thinking**:

Don't throw away old beliefs. **Update them smartly** when new clues appear.

Would you like me to visualize this with a diagram? Or show how it's used in real AI/ML examples?

Yes that would be great

Awesome! Let's visualize **Bayes' Rule** and then see how it's used in **real AI/ML examples** — all in simple language.



🎨 Visual: Bayes' Rule as a Flowchart

Imagine this simple scenario:

You want to know: Is it raining?

You look outside and see: People carrying umbrellas 🌴

Let's draw this out.

```
markdown
  [Is it raining?] ←——— (Your Question)
  [Prior Belief] (e.g. 30% chance it rains today)
[You see people with umbrellas] ← Evidence
[How likely are umbrellas if it's raining?] → 90%
[How likely are umbrellas if it's not raining?] → 20%
 Combine everything using Bayes' Rule:
 Posterior = (Prior × Likelihood) / Evidence
[Updated Belief] (Now, maybe 80% sure it's raining!)
```

You've used Bayes' Rule to update your belief smartly based on new evidence!

🔖 In Real AI/ML: Where Bayes' Rule is Used

1. Spam Detection (Naive Bayes Classifier)

Let's say:

- Prior: How likely a message is spam (e.g., 20%)
- Likelihood: How often the word "free" appears in spam messages
- Evidence: The message contains "free"

Then Bayes' Rule helps calculate:

What's the chance this message is spam given that it has "free"?

This gives:

- Posterior: P(Spam | Message = "You won free money!")
- This is exactly how spam filters work!

2. Medical Diagnosis

- Prior: Only 1% people have a rare disease
- Evidence: A patient tested positive
- Likelihood: Test is 99% accurate

Even with a positive test, Bayes' Rule might tell us:

"Hold on! The disease is still rare. There's a good chance this is a false alarm."

This avoids false positives in medical tests.

3. Robotics / Self-Driving Cars

A robot may think:

- Prior: "I think I'm at this position on the map."
- Then it sees a wall nearby.

- Likelihood: "Would I see this wall if I were at that position?"
- **⇒** Bayes' Rule helps it correct its **location belief** as it moves and sees new things.

This is called **Bayesian Localization**.

③ Summary Visual (What Bayes' Rule Does)

```
BELIEF (Prior)

+
CLUE MATCH (Likelihood)

↓

TOTAL CLUE CHANCE (Evidence)

↓

= Updated BELIEF (Posterior)
```

Would you like me to create a diagram as an image for this explanation? Or do a step-by-step Naive Bayes classification example next?

help me to understand, so Prior is also a posterior of some previous event, isn't?

Property of the state of the st

Yes — the **prior** of *today* is often the **posterior** from *yesterday*.

Let's understand this in layers:

a Think like a student:

You have a notebook. On **Day 1**, you have no clue if it rains much in June.

You check weather data \rightarrow it rained 20 out of 30 days.

You write:

"Rain in June: $20/30 \rightarrow 67\%$ chance"

This is your **posterior** after observing data.

Now on Day 2:

You treat that 67% as your prior belief.

Then today, you observe dark clouds \Rightarrow new clue! You update:

"Based on the clouds, rain looks more likely today."

→ Now your **posterior** might be 85%.

This is called Recursive Bayesian Updating

Every time:

- 1. You get new data
- 2. You update your belief (posterior)
- 3. That **becomes your new prior** for the next round!

Like this:

$$Prior_t \xrightarrow{+ Evidence_t} Posterior_t = Prior_{t+1}$$

Keal-World Analogy: Learning Over Time

Let's say you're trying to guess what snacks your friend likes:

- Day 2: You assume that again (prior now) → He chooses chocolate again → Your belief grows stronger New Posterior
- And so on...

Your belief evolves as new evidence comes in!

So in short:

Yes! A prior is just a posterior from a previous learning step. Bayesian thinking is like **stacking wisdom** — one update at a time.

Would you like me to show this in a timeline or flowchart diagram?

Why the naming is posterior? any special reason?

Great question — and yes, the names "prior" and "posterior" come from Latin roots and follow a logical reason related to time and order.

Why is it called Posterior?

"Prior" → Latin: prioris → means earlier or before

It represents:

What you believed **before** seeing the new evidence.

"Posterior" → Latin: posterioris → means later or after

It represents:

What you believe **after** incorporating new evidence.

o So, the names mean:

Term	Meaning	When it exists
Prior	Earlier belief (before clue)	Before data
Posterior	Later belief (after clue)	After data

Example with naming:

Let's say we're solving a mystery:

- Before any clues → You guess: "It's probably the butler." → That's your **prior** suspicion.
- After finding a bloody glove in the garden → You update your belief → "Now I'm more sure it's the butler." → That's your posterior suspicion.

You moved from **prior knowledge** → **posterior knowledge**.

Why it's useful:

This naming helps us track **what came before** and **what came after** in a process of learning or belief updating.

Memory Tip:

"Prior" comes Prior
"Posterior" comes Post

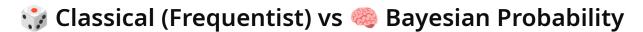
Just like in English:

- Post-test = after the test
- Post-event = after the event
- Posterior belief = belief after seeing the data

Would you like me to turn this into a visual timeline for easier memory?

Why Bayesian Probability is different than normal probability?

Excellent question! Let's break it down simply:



Both are ways to **measure uncertainty**, but they have **different philosophies**.

1. Frequentist Probability – "Long-term frequency"

- "Probability is how often something happens if you repeat it many times."
- If you flip a coin 1,000 times and get heads 500 times \rightarrow P(Heads) = 0.5
- It's about **objective reality**, not belief.
- You can't assign probability to a one-time event like:
 - "What's the chance this person has cancer?" \leftarrow Not allowed here.

📌 Example:

"There's a 1% chance of rain" \rightarrow means 1 out of 100 such days, it rained.

🧠 2. Bayesian Probability - "Degree of belief"

"Probability is a measure of how strongly I believe something is true, given what I know."

- It's **subjective**, based on current knowledge.
- You can assign probabilities to one-time events:

"Based on symptoms and past data, there's a 70% chance this person has cancer."

📌 Example:

"There's a 70% chance it'll rain tomorrow" \rightarrow means given what we know now, rain seems likely.

🔁 Updating Beliefs: Key Feature of Bayesian Probability

- Bayesians update their beliefs when new data comes in.
- This is done using Bayes' Rule.

Frequentists don't update beliefs — they wait for a lot of data and run hypothesis tests.

Summary Table

Feature	Frequentist	Bayesian
Probability of events	Long-run frequency	Degree of belief
Handles one-time events?	X No	✓ Yes
Updates with new data	X No (fixed method)	✓ Yes (via Bayes' Rule)
Example question	"How often does a fair coin land heads?"	"Given this coin, how likely is it fair?"
Uses prior knowledge?	X No	✓ Yes (via prior)

A Quick Analogy:

Frequentist: "I'll only believe it if I see it happen a thousand times."

Bayesian: "I'll start with what I know and update my belief as I learn more."

Would you like a real-world example comparing both side-by-side (like medical testing or coin guessing)?

So how to decide when to use what type of probability? I mean how the professionals choose that? Or what is the basis of choosing between these two?

Excellent question! This is exactly what professional statisticians and data scientists debate and decide based on the **problem**, **data availability**, and **philosophy of reasoning**.

1 How to Decide: Bayesian vs Frequentist

Here's a guide to **how professionals choose** between them:

Use Bayesian Probability when:

Situation	Why Bayesian Works Well
You need to update your belief as new data comes	Bayesian lets you use prior + new evidence
Data is limited or noisy	You can inject prior knowledge to stabilize learning
You care about personalized, probabilistic inference	You can get probability of a hypothesis
You want probability of a model or parameter	Bayesian inference allows this: *P(model

Situation	Why Bayesian Works Well
🧖 Medical, Risk, Forecasting	One-time, belief-based decisions: "What's the chance this person has X?"
in: Machine learning, AI agents, spam filters, robotics, forecasting	

☑ Use **Frequentist Probability** when:

Situation	Why Frequentist Works Well
You have lots of data and repeated trials	You can rely on long-run frequencies
Nou need hypothesis testing or confidence intervals	Built-in tools like p-values, t-tests, etc.
You want a "neutral", data-only approach	No subjective prior — just the data
Science experiments, clinical trials	Often relies on frequentist p-values
Simple models with known distributions	Binomial, Poisson, etc. tested over 1000s of trials

Philosophy Behind Each:

Question	Frequentist View	Bayesian View
What is probability?	Long-run frequency of events	Degree of belief, given knowledge
Can we assign probability to a hypothesis?	X No	✓ Yes
Is prior belief allowed?	X No (objective)	✓ Yes (explicit)

🔍 Real-World Examples:

Problem	Frequentist Way	Bayesian Way
Coin fairness	Flip it 1000 times and measure heads frequency	Assume prior belief (e.g., 50%) and update with each flip
Is this drug effective?	Use p-values and confidence intervals	Use prior data + trial results to get P(Effective
Predict rain tomorrow	Use historical weather frequency	Use weather model + today's satellite data

TL;DR Summary:

Choose Bayesian if	Choose Frequentist if
You have small data + expert knowledge	You have lots of repeated data
You want personalized belief updates	You want objective , repeatable results
You want probabilities of models/hypotheses	You care about estimates or tests
You need flexible modeling	You need rigorous classical tools

🧩 Final Tip:

Most modern ML and AI systems lean toward Bayesian thinking. But for scientific experiments, frequentist methods are still very dominant.

Would you like a decision flowchart or checklist to help choose which to use in your projects?

But tell me one thing, when we are calculating first posterior i.e. when no prior is available, that we are calculating using frequency only right? isn't that frequentist?

🔥 Brilliant observation — and you're absolutely right to raise this!

Let's walk through it clearly:

Or Your Question:

When we calculate the **first posterior** (when we have **no prior**), don't we just use frequencies?

Isn't that... Frequentist?

Short Answer:

Yes — when no prior is available, we often use frequentist methods (like observed data frequency) to initialize our Bayesian process.

But there's a subtle twist. Let's unpack it.

🔁 Bayesian Needs a Prior — So What if We Don't Have One?

When a Bayesian has no prior knowledge, we do one of these:

1. Use a Uniform (Flat) Prior

"All outcomes are equally likely."

Example:

For a coin, P(Heads) = P(Tails) = $0.5 \rightarrow a$ flat prior

This is **not based on frequency**, just ignorance — we assume nothing.

2. Use an Empirical (Data-Driven) Prior

We look at **data frequency** and treat that as the prior

For example:

- "We've seen 3 heads and 2 tails so far, so prior = Beta(3,2)"
- This is where Bayesian and Frequentist start to overlap.

Yes — you're borrowing frequentist ideas to **bootstrap your Bayesian model**.

This Overlap Has a Name

Empirical Bayes

- It's a method where:
- You estimate priors from data (frequentist-style)
- Then continue with Bayesian updating
- So yes, you're right: The **first step** is often based on **frequencies** and that's OK.

So, what's the difference then?

Feature	Frequentist	Bayesian (w/ Empirical Prior)
What is the "truth"?	A fixed, unknown value	A distribution over possible values
Prior used?	X No	Yes, even if estimated from data
Can we update belief?	X Not directly	Yes, with each new data point

Final Insight:

Even though Bayesians sometimes **start with frequencies**, the **moment they use those frequencies to define a prior**, they switch into a **Bayesian framework**.

So yes, you're right — the **line blurs** at the start, but the **philosophy** and **what comes next** is what makes it **Bayesian**.

Would you like to see an example showing how we go from frequency \rightarrow prior \rightarrow posterior in a Bayesian model?