Bayes’ Theorem

Emile Latour, Nicky Wakim, Meike Niederhausen

January 19, 2026

Recap: What we did last time

Last class, we calculated the **Positive Predictive Value (PPV)** using a contingency table:

	Disease	No Disease	Total
Test +	90	990	1,080
Test -	10	8,910	8,920
Total	100	9,900	10,000

$$P(\text{Disease} \mid \text{Test } +) = \frac{90}{1,080} \approx 0.083$$

Today we'll see the **formula** that lets us calculate this directly.

Bayes' Theorem

The question we're asking

We want to **flip the conditional probability**:

What we know:

- $P(\text{Test}+ \mid \text{Disease})$ (sensitivity)
- $P(\text{Test}+ \mid \text{No Disease})$ (1 - specificity)
- $P(\text{Disease})$ (prevalence)

What we want:

- $P(\text{Disease} \mid \text{Test}+)$ (PPV)

Bayes' Theorem gives us a formula to "flip" the conditional.

Bayes' Theorem: General form

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

In words:

- $P(A \mid B)$ = **posterior** (what we want to know)
- $P(B \mid A)$ = **likelihood** (what we can measure)
- $P(A)$ = **prior** (base rate)
- $P(B)$ = **marginal probability** of B

Bayes' Theorem: Medical testing version

For our diagnostic test:

$$P(\text{Disease} \mid \text{Test} +) = \frac{P(\text{Test} + \mid \text{Disease}) \times P(\text{Disease})}{P(\text{Test} +)}$$

In words:

- $P(\text{Disease} \mid \text{Test} +)$ = PPV (what we want)
- $P(\text{Test} + \mid \text{Disease})$ = sensitivity (what test manufacturer tells us)
- $P(\text{Disease})$ = prevalence (base rate in population)
- $P(\text{Test} +)$ = overall rate of positive tests

The tricky part: $P(\text{Test }+)$

How do we find $P(\text{Test }+)$?

There are **two ways** to test positive:

1. Have disease AND test positive: $P(\text{Disease and Test }+)$
2. No disease AND test positive: $P(\text{No Disease and Test }+)$

Using the **law of total probability**:

$$P(\text{Test }+) = P(\text{Test }+ \mid \text{Disease}) \times P(\text{Disease}) + P(\text{Test }+ \mid \text{No Disease}) \times P(\text{No Disease})$$

Bayes' Theorem: Complete formula

$$P(\text{Disease} \mid \text{Test} +) = \frac{P(\text{Test} + \mid \text{Disease}) \times P(\text{Disease})}{P(\text{Test} + \mid \text{Disease}) \times P(\text{Disease}) + P(\text{Test} + \mid \text{No Disease}) \times P(\text{No Disease})}$$

This looks complicated, but we have all the pieces!

Applying Bayes' Theorem to our example

Given:

- Sensitivity = 90% = $P(\text{Test} + \mid \text{Disease}) = 0.90$
- Specificity = 90%, so $P(\text{Test} + \mid \text{No Disease}) = 0.10$
- Prevalence = 1% = $P(\text{Disease}) = 0.01$
- Therefore: $P(\text{No Disease}) = 0.99$

Step 1: Calculate the numerator

$$\begin{aligned}\text{Numerator} &= P(\text{Test} + \mid \text{Disease}) \times P(\text{Disease}) \\ &= 0.90 \times 0.01 \\ &= 0.009\end{aligned}$$

This represents the **true positives** (people who have disease AND test positive).

Step 2: Calculate the denominator

$$\begin{aligned}\text{Denominator} &= P(\text{Test} + \mid \text{Disease}) \times P(\text{Disease}) \\ &\quad + P(\text{Test} + \mid \text{No Disease}) \times P(\text{No Disease}) \\ &= (0.90 \times 0.01) + (0.10 \times 0.99) \\ &= 0.009 + 0.099 \\ &= 0.108\end{aligned}$$

This represents **all positive tests** (true positives + false positives).

Step 3: Calculate PPV

$$\begin{aligned} P(\text{Disease} \mid \text{Test } +) &= \frac{0.009}{0.108} \\ &\approx 0.083 \\ &= 8.3\% \end{aligned}$$

Same answer as before! But now we calculated it directly from the formula.

Connecting to the contingency table

Bayes' Theorem:

$$\frac{0.90 \times 0.01}{(0.90 \times 0.01) + (0.10 \times 0.99)} = 0.083$$

Numerator: true positives

Denominator: all positives

The formula gives us the **proportions** directly!

Contingency table:

	Disease	No Disease	Total
Test +	90	990	1,080
Test -	10	8,910	8,920
Total	100	9,900	10,000

$$\frac{90}{1,080} = 0.083$$

Why is this useful?

Without Bayes' Theorem:

- Build entire contingency table
- Need to choose a population size
- More steps, more chances for errors

With Bayes' Theorem:

- Plug numbers directly into formula
- Get answer immediately
- Works with any probabilities (don't need counts)

Both approaches work - use whichever makes more sense to you!

What to remember about Bayes' Theorem

The formula:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

Key concepts:

- Bayes' Theorem "flips" conditional probabilities
- Always consider **base rates** (priors)
- Low prevalence → low PPV, even with accurate tests
- Can use formula OR contingency table (both work!)
- **Posterior = (Likelihood × Prior) / Evidence**

Bayes vs. Contingency tables

Contingency table approach:

Pros:

- Visual and intuitive
- See all relationships
- Good for teaching

Cons:

- Requires choosing population size
- More steps
- Can’t easily see effect of changing parameters

Bayes’ Theorem approach:

Pros:

- Direct calculation
- Works with proportions
- Easy to see how prevalence affects PPV

Cons:

- Formula can seem abstract
- Need to remember to calculate denominator

Use whichever approach makes more sense to you!