

# A Gentle Example on Bayesian Inference: Hemophilia

*Disclaimer: The following content is heavily inspired by [this video](#) and the example is taken from [this book](#). This is not intended as a full explanation of Bayesian inference, only as a visualization to help understand key ideas. That said, even though the video is awesome for learning, I think the area-based visual style (the square diagrams) doesn't quite scale to more abstract or realistic cases. So I'm trying to keep the great intuition but bring it into a format that suits other cases.*

## Introduction

**Hemophilia** is a rare genetic disorder in which the blood doesn't clot properly due to a lack of certain clotting factors. It is usually inherited (rarely mutated) in an X-linked recessive pattern, meaning it is passed down through the X chromosome from the mother.

**i** How does chromosome inheritance work?

In humans, chromosomes are inherited from both parents — 23 from the mother and 23 from the father — for a total of 46 chromosomes arranged in 23 pairs. Of these, 22 pairs are called **autosomes**, and the 23rd pair determines **biological sex**.

- Females: XX (two X chromosomes)
- Males: XY (one X and one Y chromosome)

The egg (from the mother) always contributes an X. The sperm (from the father) can contribute either an X or a Y. So: - X + X = **female** - X + Y = **male**

Since the affected gene is on the X chromosome, a male who inherits the mutation on his single X chromosome will have hemophilia. A female would need to inherit the mutation on **both** of her X chromosomes to be affected. If only one of her X chromosomes carries the mutation, she is a **carrier** — she can pass on the gene but typically shows no symptoms.

We can illustrate key concepts of **Bayesian inference** using this example.

### The Problem

A couple has two kids: one daughter and a newborn son. The son is diagnosed with hemophilia. Tests reveal the father is not hemophilic and the mother is a carrier.

20 years later, the daughter is now married and expecting a child. Her husband is not hemophilic. She wonders: *what is the probability that she is a carrier? Should she undergo testing to better inform her children?*

## Bayesian Approach: Calculations

Bayesian statistics shine when new evidence becomes available over time, allowing you to update your beliefs progressively.

Let's assume a sample space  $\Omega$ , where event  $H = \{\text{she is a carrier}\}$  has prior probability  $P(H) = 0.5$ .

When new evidence  $E$  is observed, Bayes' theorem states:

$$P(H | E) = \frac{P(E | H) \cdot P(H)}{P(E | H) \cdot P(H) + P(E | \sim H) \cdot P(\sim H)}$$

The magic here is that now we're not talking about  $P(H)$  in a vacuum—we're talking about  $P(H | E)$ , how likely  $H$  is given what we saw  $E$ . This is often depicted as restricting our sample space only to the cases where  $E$  is favorable.

In this hemophilia case, the most useful kind of evidence is whether her kids have the disease or not. We can use what we know about genetics to estimate how likely each outcome is.

- If she has a first kid, let's say  $y_1$ , we have two possibilities:
  - If he is not affected by the disease  $y_1 = 0$ :
    - \*  $P(y_1 = 0 | H) = 0.5$ . Because the boy has a 50% chance to inherit the affected X chromosome from his mother.
    - \*  $P(y_1 = 0 | \sim H) = 1$ . Because there is no chance to inherit any affected chromosome.
  - If he is affected by the disease  $y_1 = 1$ :
    - \*  $P(y_1 = 1 | H) = 0.5$ . Because the boy has a 50% chance to inherit the affected X chromosome from his mother.
    - \*  $P(y_1 = 1 | \sim H) = 0$ . Because there is no chance to inherit any affected chromosome.

- If she has two kids,  $y_1$  and  $y_2$ , then our possibilities look like this:
  - Both are not affected by the disease ( $y_1 = 0, y_2 = 0$ ):
    - \*  $P(y_1 = 0, y_2 = 0 \mid H) = 0.5 \times 0.5 = 0.25$ : Because each kid independently has a 50% chance of inheriting the unaffected X chromosome from a carrier mother.
    - \*  $P(y_1 = 0, y_2 = 0 \mid \sim H) = 1 \times 1 = 1$ : Because if the mother is not a carrier, neither child can inherit an affected chromosome.
  - First one is not affected but the second is ( $y_1 = 0, y_2 = 1$ ):
    - \*  $P(y_1 = 0, y_2 = 1 \mid H) = 0.5 \times 0.5 = 0.25$ : Because each kid independently has a 50% chance of inheriting the unaffected X chromosome from a carrier mother.
    - \*  $P(y_1 = 0, y_2 = 1 \mid \sim H) = 0.5 \times 0 = 0$ : Because it is (nearly) impossible for the second kid to be affected if his mum is not affected.
  - The case ( $y_1 = 1, y_2 = 0$ ) is homologous.
  - Both kids are affected, this is ( $y_1 = 1, y_2 = 1$ ):
    - \*  $P(y_1 = 1, y_2 = 1 \mid H) = 0.5 \times 0.5 = 0.25$ : Because each kid independently has a 50% chance of inheriting the unaffected X chromosome from a carrier mother.
    - \*  $P(y_1 = 1, y_2 = 1 \mid \sim H) = 0$ : Because it is (nearly) impossible for the second kid to be affected if his mum is not affected.

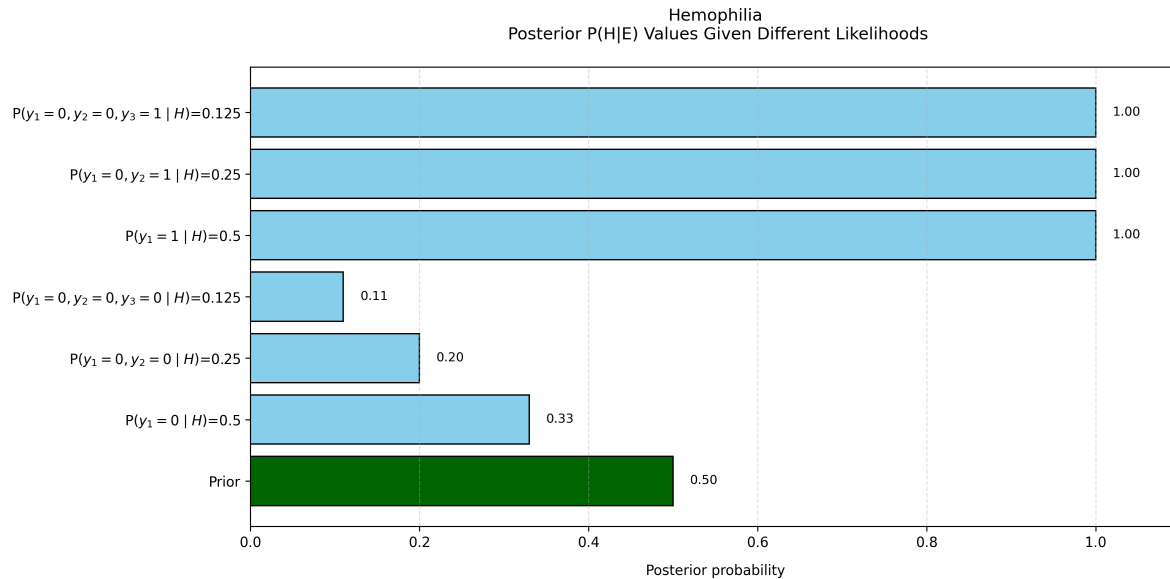
*Note how the use of the word likely is not accidental, since this part of the equation is often called **likelihood**. Typically the likelihood function is selected given the data. Here we are omitting the step of drawing from a binomial because the case is so simple and we take it for granted.*

Solving the Bayes equation for this likelihoods we have the following results:

$$\begin{aligned}
 P(H \mid y_1 = 0) &= \frac{P(y_1 = 0 \mid H) \cdot P(H)}{P(y_1 = 0 \mid H) \cdot P(H) + P(y_1 = 0 \mid \sim H) \cdot P(\sim H)} \\
 &= \frac{0.5 \cdot 0.5}{0.5 \cdot 0.5 + 1 \cdot 0.5} \approx 0.33 \\
 P(H \mid y_1 = 1) &= \frac{P(y_1 = 1 \mid H) \cdot P(H)}{P(y_1 = 1 \mid H) \cdot P(H) + P(y_1 = 1 \mid \sim H) \cdot P(\sim H)} \\
 &= \frac{0.5 \cdot 0.5}{0.5 \cdot 0.5 + 0 \cdot 0.5} = 1 \\
 P(H \mid y_1 = 0, y_2 = 0) &= \frac{P(y_1 = 0, y_2 = 0 \mid H) \cdot P(H)}{P(y_1 = 0, y_2 = 0 \mid H) \cdot P(H) + P(y_1 = 0, y_2 = 0 \mid \sim H) \cdot P(\sim H)} \\
 &= \frac{0.25 \cdot 0.5}{0.25 \cdot 0.5 + 1 \cdot 0.5} = 0.20 \\
 P(H \mid y_1 = 0, y_2 = 1) &= \frac{P(y_1 = 0, y_2 = 1 \mid H) \cdot P(H)}{P(y_1 = 0, y_2 = 1 \mid H) \cdot P(H) + P(y_1 = 0, y_2 = 1 \mid \sim H) \cdot P(\sim H)} \\
 &= \frac{0.25 \cdot 0.5}{0.25 \cdot 0.5 + 0 \cdot 0.5} = 1
 \end{aligned}$$

## Bayesian Approach: Visualizations

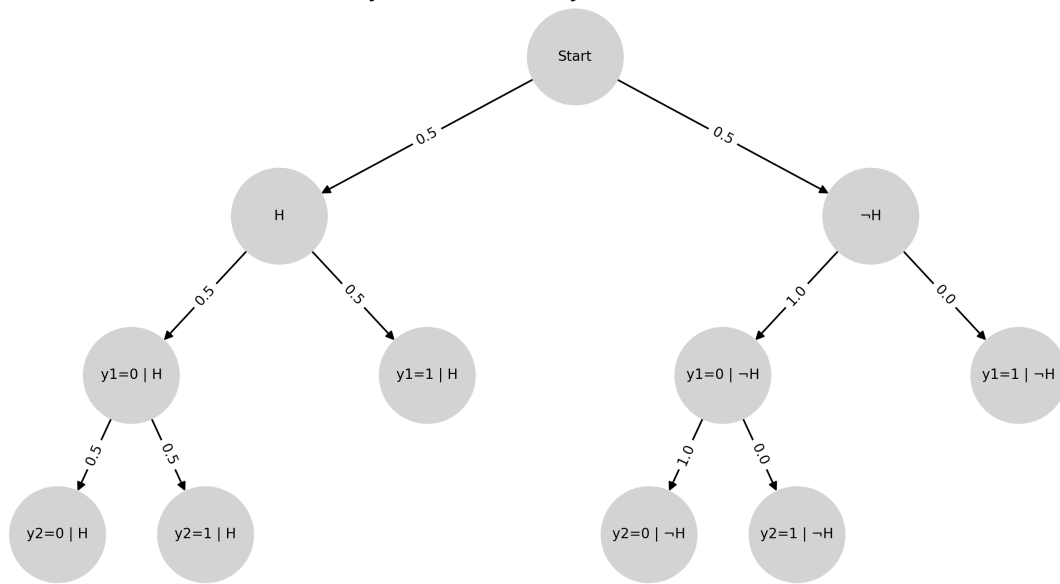
To get a better sense of how our beliefs shift, let's plot what happens when we plug in different outcomes.



This chart shows how the posterior probability changes based on different combinations of kids and outcomes. You can see how quickly our belief updates as we observe more data.

Another way to break it down is with a good old-fashioned probability tree. These are super helpful when reasoning through sequential events and can make conditional logic more visual and digestible.

Bayesian Probability Tree: 2 Sons



This tree maps out how each child's outcome splits from the previous one, letting you trace how probabilities evolve step by step. Super intuitive when you want to make sure your reasoning checks out.