Exercises on Bayesian Modeling and Reasoning

Vasilis Gkolemis

June 2025

Instructions These four exercises are designed to help you practice Bayesian reasoning using real-world style problems. They should take approximately 30 minutes to complete in total. Show your calculations clearly and write a brief explanation for each answer.

1   Exercise: Medical Test Puzzle

A rare disease affects 1% of the population. A test for the disease has:

- Sensitivity (true positive rate), i.e., the probability of testing positive given that you have the disease: 99%.

- Specificity (true negative rate), i.e., the probability of testing negative given that you do not have the disease: 95%.

If a randomly selected person tests positive, what is the probability that they actually have the disease? Before making your calculations, make an intuitive guess about the answer. Whould you expect it to be over or under 50%? Afterwards, compare your guess with the result of your calculations. Do you think the result is surprising? Why or why not?

Solution.   Let $D = \{0, 1\}$ be the event that a person {has not, has} the disease, and $T = \{0, 1\}$ be the event that a person tests {negative, positive}. We want to find: $P(D = 1|T = 1)$. We know:

- $P(D = 1) = 0.01$ (prevalence of the disease)

- $P(D = 0) = 0.99$ (no disease)

- $P(T = 1|D = 1) = 0.99$ (sensitivity)

- $P(T = 0|D = 0) = 0.95$ (specificity)

- $P(T = 1|D = 0) = 1 - P(T = 0|D = 0) = 0.05$ (false positive rate)

- $P(T = 0|D = 1) = 1 - P(T = 1|D = 1) = 0.01$ (false negative rate)

We apply Bayes' theorem:

$$P(D = 1|T = 1) = \frac{P(T = 1|D = 1)P(D = 1)}{P(T = 1)}$$

We compute $P(T)$ via the law of total probability:

$$P(T = 1) = P(T = 1, D = 1) + P(T = 1, D = 0)$$
$$= P(T = 1|D = 1)P(D = 1) + P(T = 1|D = 0)P(D = 0)$$
$$= 0.99 \cdot 0.01 + (1 - 0.95) \cdot 0.99$$
$$= 0.99 \cdot 0.01 + 0.05 \cdot 0.99$$
$$= 0.0099 + 0.0495$$
$$= 0.0594$$

Therefore:

$$P(D = 1|T = 1) = \frac{P(T = 1|D = 1)P(D = 1)}{P(T = 1)}$$
$$= \frac{0.99 \cdot 0.01}{0.0594}$$
$$= 0.99 \cdot 0.1684$$
$$\approx 0.1667$$

Answer: $\boxed{\approx 16.7\%}$

Intuitive Guess: Most people would expect this probability to be over 50%, because the test is highly sensitive. However, due to the rarity of the disease and the false positive rate, the actual probability is much lower than expected. This illustrates the importance of considering base rates in Bayesian reasoning, as the rarity of the disease significantly affects the probability even with a highly sensitive test.

2 Exercise: The Two-Card Problem

You have two cards (i) One card is red on both sides (RR) and (ii) One card is red on one side and black on the other (RB). You pick a card at random and place it on the table. The side facing up is red.

Question: What is the probability that the other side of the card is also red?

Solution. We enumerate all the red sides that could be showing up:

- The RR card has two red sides, so it can appear as either side: RR1 or RR2.

- The RB card has one red side: RB-red.

So the possible ways a red side can face up are: RR1, RR2, RB-red. In two of them (RR1 and RR2), the hidden side is also red. In one of them (RB-red), the hidden side is black. So, the probability that the other side is red given a red side is up is:

$$P(\text{other side is red} \mid \text{red side is up}) = \frac{\text{Number of favorable outcomes}}{\text{Total outcomes}}$$
$$= \frac{2}{3}$$

Answer: $\boxed{2/3}$

Intuitive Guess: Many people guess 1/2, reasoning that since there are two cards, and both are possible to be picked since both have a red side, the chances are even. However, although both cards have a red side, the RR is more likely to be picked when a red side is seen because it has two red sides. Specifically, it has a $\frac{2}{3}$ chance to be picked, while the RB card has only a $\frac{1}{3}$ chance to be picked.

3 Exercise: The Goalkeeper Problem

A football manager compares two goalkeepers, goalkeeper A and goalkeeper B, based on their save percentages against penalty shooters.

- Goalkeeper A has a better save percentage than goalkeeper B when the penalty shooter is **right-footed**.

- Goalkeeper A has a worse save percentage than goalkeeper B when the penalty shooter is **left-footed**.

Question: Is it possible for goalkeeper B to have a higher overall save percentage than goalkeeper A?

Before calculating or reasoning, make an intuitive guess: Does it seem reasonable that someone worse in both subgroups could be better overall? Afterwards, compare your intuition with the result.

Solution. Yes, it is possible for goalkeeper B to have a higher overall save percentage than goalkeeper A. This is an example of Simpson's Paradox, where trends that appear in separate groups can reverse when the groups are combined. Let's construct a hypothetical example. Suppose:

Right-footed shooters:

- Goalkeeper A: saved 8 out of 10 (80%)

- Goalkeeper B: saved 45 out of 60 (75%)

Left-footed shooters:

- Goalkeeper A: saved 30 out of 60 (50%)

- Goalkeeper B: saved 1 out of 3 (33%)

Overall:

- Goalkeeper A: saved 38 out of 70 (54.3%)

- Goalkeeper B: saved 46 out of 63 (73%)

Even though A outperforms B against right-footed shooters (80% vs 75%) and underperforms against left-footed shooters (10% vs 75%), B ends up with a significantly higher overall save percentage. This happens because the number of shots from each footedness differs significantly between goalkeepers, creating a weighted imbalance.

Answer: $\boxed{\text{Yes, due to Simpson's Paradox}}$

Intuitive Guess: Many would think that being better in both subgroups guarantees a better overall result. However, this example reveals how the composition of the subgroups (i.e., how many shots come from each type of shooter) can skew the overall result. This demonstrates the importance of looking beyond subgroup analysis and understanding how aggregated statistics can be misleading.

4 Exercise: The Monty Hall Problem

You are on a game show and presented with three doors: behind one door is a car (the prize you want), and behind the other two doors are goats (which you do not want). You choose one door, say Door 1. The host, who knows what is behind each door, opens another door, say Door 3, which has a goat behind it. Then, the host gives you the option to either stick with your original choice (Door 1) or switch to the remaining unopened door (Door 2). **Question:** Should you stick with your original choice or switch to the other door? What is the probability of winning the car if you switch versus if you stick?

Solution. When you initially pick one of the three doors (say Door 1), the probability that the car is behind that door is $\frac{1}{3}$. The probability that the car is behind one of the other two doors is $\frac{2}{3}$.

Now the host opens a door (say Door 3) and reveals a goat. This action is not random: the host always opens a door with a goat, never the one with the car. This effectively concentrates the entire $\frac{2}{3}$ probability on the remaining unopened door (Door 2), because Door 3 has been revealed as a goat.

So:

- If you **stick** with your original choice (Door 1), your chance of winning is still $\frac{1}{3}$.

- If you **switch** to the other unopened door (Door 2), your chance of winning is $\frac{2}{3}$.

Therefore, you should **switch** doors.

Answer: Switch; probability of winning if you switch is 2/3, if you stay it's 1/3

Intuitive Guess: Many people guess that the odds are 1/2 because two doors remain. However, this overlooks the fact that the host's action gives you information. Since the host always opens a goat door, switching gives you the advantage of the original 2/3 chance that the car was not behind your first pick.

Detailed Explanation: Let's enumerate all possible cases to understand why switching increases your chance of winning. There are 3 doors: Door 1, Door 2, Door 3. Assume you initially pick Door 1. The same logic applies regardless of which door you pick first, due to symmetry. The car can be behind any door with equal probability $\frac{1}{3}$:

- **Case 1:** Car behind Door 1 (your initial pick)

    - Host opens Door 2 or Door 3 (both have goats)
    - If you stick with Door 1, you win (car)
    - If you switch, you lose

- **Case 2:** Car behind Door 2

    - Host must open Door 3 (goat)
    - If you stick with Door 1, you lose
    - If you switch to Door 2, you win (car)

- **Case 3:** Car behind Door 3

    - Host must open Door 2 (goat)
    - If you stick with Door 1, you lose
    - If you switch to Door 3, you win (car)

Summary:

— Scenario — Initial Choice — Host Opens — Stick Win? — Switch Win? — ——————— ———————————-—————————————————————-— — Car behind Door 1 — Door 1 — Door 2 or 3— Yes — No — — Car behind Door 2 — Door 1 — Door 3 — No — Yes — — Car behind Door 3 — Door 1 — Door 2 — No — Yes —

Since the car is equally likely behind each door, each scenario has probability $\frac{1}{3}$.

- Probability of winning if you stick $= \frac{1}{3}$ - Probability of winning if you switch $= \frac{2}{3}$

Intuitive Guess: While it may seem that after a door is opened the chance is 50-50 between the two remaining doors, this table shows the host's knowledge changes the probabilities. Switching leverages the host's action, effectively transferring the initial $\frac{2}{3}$ probability of the car not being behind your first pick to the other unopened door.

5 Exercise: Blue Cab Problem

In a city, 85% of the taxis are green and 15% are blue. A hit-and-run accident occurs at night. A witness identifies the cab involved as blue. The witness is 80% reliable in distinguishing blue from green in such conditions. **Question:** What is the probability that the cab involved in the accident was actually blue?

Solution. We define events:

- B: the cab involved is blue.

- G: the cab involved is green.

- W_B: witness says the cab is blue.

- W_G: witness says the cab is green.

Given:
$$P(B) = 0.15, \quad P(G) = 0.85$$

Witness reliability:

- $P(W_B \mid B) = 0.80$ (witness correctly identifies blue)

- $P(W_G \mid G) = 0.80$ (witness correctly identifies green)

Therefore, the witness error rates are:

- $P(W_B \mid G) = 1 - 0.80 = 0.20$ (incorrectly says blue when green)

- $P(W_G \mid B) = 1 - 0.80 = 0.20$ (incorrectly says green when blue)

We want to find the probability that the cab was actually blue given the witness said blue:

$$P(B \mid W_B) = ?$$

By Bayes' theorem:
$$P(B \mid W_B) = \frac{P(W_B \mid B)P(B)}{P(W_B)}$$

We compute $P(W_B)$ by total probability:

$$P(W_B) = P(W_B \mid B)P(B) + P(W_B \mid G)P(G)$$
$$= 0.80 \times 0.15 + 0.20 \times 0.85$$
$$= 0.12 + 0.17$$
$$= 0.29$$

So:

$$P(B \mid W_B) = \frac{0.80 \times 0.15}{0.29} = \frac{0.12}{0.29} \approx 0.4138$$

Answer: $\boxed{\approx 41.4\%}$

Intuitive Guess: Although the witness is fairly reliable (80%), because blue cabs are much rarer (15%), many of the "blue" reports are actually mistakes when the cab was green. Hence, the actual probability that the cab was blue given the witness says blue is only about 41%, less than half.