Tutoriat PS 8: Bayes actualization with discrete a priori

Mihai Duzi, Lucan Cristian, Luminaru Ionut

December 10, 2025

Contents

1 Bayes a priori/a posteriori probabilities

1.1 Example: Rigged coins

There is three types of coins:
A with a probability of heads of 0.5
B with a probability of heads of 0.6
C with a proability of heads of 0.9

Presume we have a drawer with 5 coins: 2 of type A, 2 of type B and one of type C. We randomly pick a coin from the drawer. Without looking at it, we toss it once and obtain heads. What is the probability that it was of type A? Of type B? Of type C?

Answer.

Let A, B, and C be the events that the chosen coin was of type A, type B, and type C, respectively. Let D be the event that the toss resulted in heads. The problem asks for

$$P(A \mid D), \qquad P(B \mid D), \qquad P(C \mid D).$$

Terminology

Experiment: We randomly pick a coin from the drawer, toss it, and record the result.

Data: the result of our experiment. In this case the event

$$D = \text{“heads”}.$$

We think of D as data that gives evidence in favor of or against each hypothesis.

Hypotheses: we test 3 hypotheses: the coin is type A, B, or C.

Prior probabilities: the probability of each hypothesis before the coin toss (before collecting data). Since the drawer has 2 coins of type A, 2 of type B, and one coin of type C, we have

$$P(A) = 0.4, \qquad P(B) = 0.4, \qquad P(C) = 0.2.$$

Likelihood

(This is the same likelihood function that we have used for MLE.) The likelihood function is $P(D \mid H)$, i.e., the probability of the data assuming the hypothesis is true. Most often we consider the data to be fixed and let the hypotheses vary. For example, $P(D \mid A)$ is the probability of getting heads if the coin is type A. In our case the likelihoods are

$$P(D \mid A) = 0.5, \qquad P(D \mid B) = 0.6, \qquad P(D \mid C) = 0.9.$$

Posterior probabilities: the probability of each hypothesis given the data obtained from the coin toss:

$$P(A \mid D), \qquad P(B \mid D), \qquad P(C \mid D).$$

Bayes' theorem says, for example,

$$P(A \mid D) = \frac{P(D \mid A)P(A)}{P(D)}.$$

The denominator $P(D)$ is computed using the law of total probability:

$$P(D) = P(D \mid A)P(A) + P(D \mid B)P(B) + P(D \mid C)P(C) = 0.5 \cdot 0.4 + 0.6 \cdot 0.4 + 0.9 \cdot 0.2 = 0.62.$$

Now each of the three posterior probabilities can be computed:

$$P(A \mid D) = \frac{P(D \mid A)P(A)}{P(D)} = \frac{0.5 \cdot 0.4}{0.62} = \frac{0.2}{0.62} \approx 0.3226,$$

$$P(B \mid D) = \frac{P(D \mid B)P(B)}{P(D)} = \frac{0.6 \cdot 0.4}{0.62} = \frac{0.24}{0.62} \approx 0.3871,$$

$$P(C \mid D) = \frac{P(D \mid C)P(C)}{P(D)} = \frac{0.9 \cdot 0.2}{0.62} = \frac{0.18}{0.62} \approx 0.2903.$$

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\vert\mathcal{H})$	$P(\mathcal{D}\vert\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\vert\mathcal{D})$
A	0.4	0.5	0.2	0.3226
B	0.4	0.6	0.24	0.3871
C	0.2	0.9	0.18	0.2903
total	1		0.62	1

Figure 1: Table representation of above exercise

Note: Summing the numerators gives P(D) (total probability law)
Note: Dividing each numerator by the whole sum gives the posterior probability

1.2 Probability mass functions a priori and a posteriori

Another way of representing the same table, but with different notations is this:

Hypothesis	θ	prior pmf $p(\theta)$	poster pmf $p(\theta\vert x=1)$
A	0.5	$P(A) = p(0.5) = 0.4$	$P(A\vert\mathcal{D}) = p(0.5\vert x=1) = 0.3226$
B	0.6	$P(B) = p(0.6) = 0.4$	$P(B\vert\mathcal{D}) = p(0.6\vert x=1) = 0.3871$
C	0.9	$P(C) = p(0.9) = 0.2$	$P(C\vert\mathcal{D}) = p(0.9\vert x=1) = 0.2903$

Figure 2: Table representation of above exercise with different notation

Where the notations are as follows:
Standard notation:
θ is the value of the hypothesis.
$p(\theta)$ is the prior probability mass function of the hypothesis.
$p(\theta \mid D)$ is the posterior probability mass function of the hypothesis given the data.
$p(D \mid \theta)$ is the likelihood function. (This is not a pmf!)

In Example 1 we can represent the 3 hypotheses A, B, and C by $\theta = 0.5, 0.6, 0.9$. For the data, $x = 1$ means heads and $x = 0$ means tails. Then the prior and posterior probabilities in the table define the prior and posterior probability mass functions. ("poster" = "posterior")

1.2.1 Coin example using tails:

This is how the table looks if we re-do the exercise but instead of heads we get tails:

hypothesis	prior	likelihood	Bayes numerator	posterior
θ	$p(\theta)$	$p(x=0\,\vert\,\theta)$	$p(x=0\,\vert\,\theta)p(\theta)$	$p(\theta\,\vert\,x=0)$
0.5	0.4	0.5	0.2	0.5263
0.6	0.4	0.4	0.16	0.4211
0.9	0.2	0.1	0.02	0.0526
total	1		0.38	1

Figure 3: Table representation of the tails version of the exercise

1.3 Coin example extended

We pick a coin from the drawer, throw it twice and get heads. What is the probability of the coin being of type A, B or C?

Solution

Now the problem asks us

$$P(A \mid H_1, H_2), \qquad P(B \mid H_1, H_2), \qquad P(C \mid H_1, H_2)$$

Using Bayes again

$$P(A \mid H_1, H_2) = \frac{P(H_1, H_2 \mid A)P(A)}{P(H_1, H_2)}, \qquad P(B \mid H_1, H_2) = \frac{P(H_1, H_2 \mid B)P(B)}{P(H_1, H_2)}, \qquad P(C \mid H_1, H_2) = \frac{P(I}{}$$

The throws are independent so

$$P(H_1, H_2 \mid A) = P(H_1 \mid A)P(H_2 \mid A) = 0.5^2 = 0.25,$$

$$P(H_1, H_2 \mid B) = P(H_1 \mid B)P(H_2 \mid B) = 0.6^2 = 0.36,$$
$$P(H_1, H_2 \mid C) = P(H_1 \mid C)P(H_2 \mid C) = 0.9^2 = 0.81.$$

With the law of total probability we get:

$$P(H_1, H_2) = P(H_1, H_2 \mid A)P(A) + P(H_1, H_2 \mid B)P(B) + P(H_1, H_2 \mid C)P(C)$$

$$= 0.25 \cdot 0.4 + 0.36 \cdot 0.4 + 0.81 \cdot 0.2 = 0.1 + 0.144 + 0.162 = 0.406.$$

And so the final results are:

$$P(A \mid H_1, H_2) = \frac{0.25 \cdot 0.4}{0.406} = \frac{0.10}{0.406} \approx 0.2463,$$

$$P(B \mid H_1, H_2) = \frac{0.36 \cdot 0.4}{0.406} = \frac{0.144}{0.406} \approx 0.3542,$$
$$P(C \mid H_1, H_2) = \frac{0.81 \cdot 0.2}{0.406} = \frac{0.162}{0.406} \approx 0.3995.$$

And the table representation using the probability mass function notation from earlier:

hypothesis	prior	likelihood 1	Bayes numerator 1	likelihood 2	Bayes numerator 2	posterior 2
θ	$p(\theta)$	$p(x_1 = 1 \mid \theta)$	$p(x_1 = 1 \mid \theta)p(\theta)$	$p(x_2 = 1 \mid \theta)$	$p(x_2 = 1 \mid \theta)p(x_1 = 1 \mid \theta)p(\theta)$	$p(\theta \mid x_1 = 1, x_2 = 1)$
0.5	0.4	0.5	0.2	0.5	0.1	0.2463
0.6	0.4	0.6	0.24	0.6	0.144	0.3547
0.9	0.2	0.9	0.18	0.9	0.162	0.3990
total	1				0.406	1

Figure 4: Table representation of the 2 heads exercise variation

When we compare two events, we can express probabilistic statements in terms of odds.

2 Odds

2.1 Definition.

The *odds* of an event E versus an event E' are the ratio of their probabilities $P(E)/P(E')$. If the event E' is not specified, the second event is assumed to be the complementary event E^c. Thus, the odds of E are:

$$O(E) = \frac{P(E)}{P(E^c)}.$$

For example, $O(\text{rain}) = 2$ means that the probability of rain is twice the probability of no rain (2/3 versus 1/3). We can say that "the odds of rain are 2 to 1."

2.2 Example: Marfan syndrome

Marfan syndrome is a genetic disorder of the connective tissue that occurs in 1 out of 15,000 people. The main ocular characteristics of Marfan syndrome include bilateral lens dislocation, myopia, and retinal detachment. Approximately 70% of people with Marfan syndrome have at least one of these ocular characteristics, while only 7% of people without Marfan syndrome have them. (We do not guarantee the accuracy of these numbers.)

If a person has at least one of these ocular characteristics, what are the odds that they have Marfan syndrome?

Answer.

This is a standard Bayesian updating problem. Our hypotheses are:

M = "the person has Marfan syndrome", M^c = "the person does not have Marfan syndrome".

The data is:

F = "the person has at least one ocular characteristic".

We know the prior probability of M and the likelihoods of F given M or M^c:

$$P(M) = \frac{1}{15000}, \qquad P(F \mid M) = 0.7, \qquad P(F \mid M^c) = 0.07.$$

We can compute the posterior probabilities using a table:

hypothesis	$P(H)$	$P(F \mid H)$	$P(F \mid H)P(H)$	$P(H \mid F)$
M	0.000067	0.7	0.0000467	0.00066
M^c	0.999933	0.07	0.069995	0.99933
total	1		0.07004	1

The prior odds are:

$$O(M) = \frac{P(M)}{P(M^c)} = \frac{\frac{1}{15000}}{\frac{14999}{15000}} = \frac{1}{14999} \approx 0.000067.$$

The posterior odds are given by the ratio of the posterior probabilities, or by the ratio of the Bayes numerators (since the normalizing factor is the same for numerator and denominator):

$$O(M \mid F) = \frac{P(M \mid F)}{P(M^c \mid F)} = \frac{P(F \mid M)P(M)}{P(F \mid M^c)P(M^c)} = 0.000667.$$

2.3 Bayes Factor

2.3.1 Definition

The Bayes factor formula where D is our data and H our hypothesis is:

$$BF = \frac{P(D \mid H)}{P(D^c \mid H)}$$

Let us see exactly where the Bayes factor appears in the updating of odds. We have

$$O(H|D) = \frac{P(H|D)}{P(H^c|D)} = \frac{P(D|H)P(H)}{P(D|H^c)P(H^c)} = \frac{P(D|H)}{P(D|H^c)} \cdot \frac{P(H)}{P(H^c)} = \frac{P(D|H)}{P(D|H^c)} \cdot O(H).$$

posterior odds = **Bayes factor** $\times$ prior odds.

From this formula we see that the Bayes factor (BF) tells us whether the data give evidence for or against the hypothesis. If $BF > 1$, then the posterior odds are larger than the prior odds. Thus, the data give evidence for the hypothesis. If $BF < 1$, then the posterior odds are smaller than the prior odds. Thus, the data give evidence against the hypothesis. If $BF = 1$, then the prior odds and the posterior odds are equal. Thus, the data do not give any evidence for or against the hypothesis.

2.3.2 Bloody example

Two people left blood traces at the scene of a crime. A suspect, Oliver, has blood type "0". At the crime scene there are two blood types: "0" (a type common in the local population, with frequency 60%) and "AB" (a rare type, with frequency 1%). Do these data (blood types "0" and "AB" found at the crime scene) provide evidence in favor of the proposition that "Oliver was one of the two persons present at the crime scene"?

Answer:

There are 2 hypotheses:

$$S = \text{"Oliver and another unknown person were at the crime scene"};$$

$$S^c = \text{"2 unknown persons were at the crime scene"}.$$

The data are:

$$D = \text{"blood types '0' and 'AB' were found"}.$$

The Bayes factor for the presence of Oliver is $BF_{\text{Oliver}} = \frac{P(D|S)}{P(D|S^c)}$. We compute numerator and denominator separately.

The data say that both blood types “0” and “AB” were found at the crime scene. If Oliver was at the scene, type “0” is already present. Therefore, $P(D|S)$ is the probability that the other person had type “AB”. We are told that this probability is 0.01, so $P(D|S) = 0.01$.

If Oliver was not at the crime scene, then there are two random persons, one with type “0” and one with type “AB”. The probability of this fact is $0.6 \cdot 0.01$. The factor 2 is because there are two ways it can happen: first person type 0 and the second type AB, or vice versa.

Thus the Bayes factor for Oliver’s presence is

$$BF_{\text{Oliver}} = \frac{P(D|S)}{P(D|S^c)} \approx \frac{0.01}{2 \cdot 0.6 \cdot 0.01} = 0.8$$

2.3.3 Bloody example extended

Another suspect, Alberto, has blood type AB. Do the same data provide evidence in favor of the proposition that Alberto was one of the two persons present at the scene?

Answer:

Reusing the notation above but with Alberto instead of Oliver, we have:

$$BF_{\text{Alberto}} = \frac{P(D|S)}{P(D|S^c)} \approx \frac{0.6}{2 \cdot 0.6 \cdot 0.01} = 50.$$

Since $BF_{\text{Alberto}} \gg 1$, the data give strong evidence in favor of Alberto’s presence at the crime scene.