FOUNDATION YEAR 2025-2026

COURSE - DATA ANALYSIS

Lesson 9

Niccolò Maffezzoli, Ca' Foscari University of Venice

February 23, 2026

Conditional probability

Conditional probability

$P(A|B)$ **is the probability of A given that B is true.**

Example 1: the probability that any given person has a cough on any given day may be only 5%. But if we know or assume that the person is sick, then they are much more likely to be coughing. For example, the conditional probability that someone who is sick is coughing might be 75%, in which case we would have that $P(Cough) = 5\%$ and $P(Cough|Sick) = 75\%$.

Example 2: if a day is chosen at random than $P(Monday) = 1/7$. However, if we know that it is a weekday, this becomes $P(Monday|weekday) = 1/5$.

- If $P(A|B) = P(A)$, then events A and B are said to be independent: in such a case, knowledge about either event does not alter the likelihood of each other.

Bayes Theorem

$$P(A|B)P(B) = P(A \text{ and } B) = P(B|A)P(A) \tag{1}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

- $P(A|B)$ is a conditional probability: the probability of event A occurring given that B is true. It is also called the posterior probability of A given B.
- $P(B|A)$ is also a conditional probability: the probability of event B occurring given that A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B respectively without any given conditions; they are known as the prior probability and marginal probability.
- IMPORTANT: Often we can find $P(B) = \sum_n P(B|A_n)P(A_n)$

Example 1: Dangerous Fires

What is the probability of dangerous Fire when there is Smoke ? Given the following:

- Dangerous fires are rare (1%), i.e. $P(Fire) = 0.01$
- Smoke is fairly common (10%), i.e. due to barbecues. $P(Smoke) = 0.1$
- We know that 90% of dangerous fires make smoke. $P(Smoke|Fire) = 0.9$

$$P(Fire|Smoke) = \frac{P(Smoke|Fire)P(Fire)}{P(Smoke)} = \frac{0.9 \cdot 0.01}{0.1} = 0.09 = 9\%$$

It is worth checking out the smoke !

Example 2: Picnic

You are planning a picnic today, but the morning is cloudy. What is the probability that it will rain ? Given the following:

- 50% of all rainy days start off cloudy!, i.e. $P(cloudy|rain) = 0.5$
- But cloudy mornings are common (about 40% of days start cloudy), i.e. $P(cloudy) = 0.4$
- And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%), i.e. $P(rain) = 0.1$

$$P(rain|cloudy) = \frac{P(cloudy|rain)P(rain)}{P(cloudy)} = \frac{0.5 \cdot 0.1}{0.4} = 0.125 = 12.5\%$$

Not too bad, let's have a picnic!

Example 3: Cat Allergy

Hunter says she is itchy. There is a test for Allergy to Cats, but this test is not always right:

- For people that really do have the allergy, the test says "Yes" 80% of the time
- For people that do not have the allergy, the test says "Yes" 10% of the time (known as "false positives")
- 1% of the population have the allergy

Hunter's test says "Yes", what are the chances that Hunter really has the allergy?

$$\begin{aligned} P(allergy|yes) &= \frac{P(yes|allergy)P(allergy)}{P(yes)} \\ &= \frac{P(yes|allergy)P(allergy)}{P(yes|allergy)P(allergy) + P(yes|nonallergy)P(nonallergy)} \\ &= \frac{0.8 \cdot 0.01}{0.8 \cdot 0.01 + 0.1 \cdot 0.99} \approx 0.075 = 7.5\% \end{aligned}$$

To calculate $P(yes)$ we have applied the formula $P(B) = \sum_n P(B|A_n)P(A_n)$
$\rightarrow P(yes) = P(yes|allergy)P(allergy) + P(yes|nonallergy)P(nonallergy)$

Example 4: The Art Competition

The Art Competition has entries from three painters: Pam, Pia and Pablo.

- Pam put in 15 paintings, 4% of her works have won First Prize.
- Pia put in 5 paintings, 6% of her works have won First Prize.
- Pablo put in 10 paintings, 3% of his works have won First Prize.

What is the chance that Pam will win First Prize?

$$P(Pam|1st) = \frac{P(1st|Pam)P(Pam)}{P(1st)}$$

$$= \frac{P(1st|Pam)P(Pam)}{P(1st|Pam)P(Pam) + P(1st|Pia)P(Pia) + P(1st|Pablo)P(Pablo)}$$

$$= \frac{4\% \cdot 50\%}{4\% \cdot 50\% + 6\% \cdot 17\% + 3\% \cdot 33\%} = 0.5$$

To calculate $P(1st)$ we have again applied the formula $P(B) = \sum_n P(B|A_n)P(A_n)$. In this example we have n=3 cases (instead of 2 of the Cat example).

Example 5: Machine defects

A factory produces screws using three machines, A1, A2 and A3:

- Machine A1 produces 50% of the screws with a probability of defect D of 1%.
- Machine A2 produces 30% of the screws with a probability of defect D of 2%.
- Machine A3 produces 20% of the screws with a probability of defect D of 5%.

where D means "the screw is defective". A screw is chosen at random and is found to be defective. What is the probability that it was produced by machine A3, i.e. $P(A3|D)$?

$$\begin{aligned} P(A3|D) &= \frac{P(D|A3)P(A3)}{P(D)} \\ &= \frac{P(D|A3)P(A3)}{P(D|A1)P(A1) + P(D|A2)P(A2) + P(D|A3)P(A3)} \\ &= \frac{5\% \cdot 20\%}{1\% \cdot 50\% + 2\% \cdot 30\% + 5\% \cdot 20\%} = 48\% \end{aligned}$$

Example 6: Airport Security Scanner

At an airport, passengers pass through a scanner. There are three types of passengers:

- A1: regular passengers (85%)
- A2: Passengers carrying metallic objects (10%)
- A3: Security-risk passengers (5%)

The scanner triggers an alarm B with different probabilities:

- $P(B|A1) = 4\%$
- $P(B|A2) = 60\%$
- $P(B|A3) = 90\%$

A randomly selected passenger triggers the alarm. What is the probability that this passenger belongs to group $P(A3|B)$ (security-risk)?

Example 6: Airport Security Scanner

At an airport, passengers pass through a scanner. There are three types of passengers:

- A1: regular passengers (85%)
- A2: Passengers carrying metallic objects (10%)
- A3: Security-risk passengers (5%)

The scanner triggers an alarm B with different probabilities:

- $P(B|A1) = 4\%$
- $P(B|A2) = 60\%$
- $P(B|A3) = 90\%$

A randomly selected passenger triggers the alarm. What is the probability that this passenger belongs to group $P(A3|B)$ (security-risk)?

$$\begin{aligned} P(A3|B) &= \frac{P(B|A3)P(A3)}{P(B|A1)P(A1) + P(B|A2)P(A2) + P(B|A3)P(A3)} \\ &= \frac{90\% \cdot 5\%}{4\% \cdot 85\% + 60\% \cdot 10\% + 90\% \cdot 5\%} = 32\% \end{aligned}$$

Example 7: Email Spam Filter

An email system automatically classifies emails as "Spam".

- 30% of incoming emails are actually spam.
- The filter correctly detects spam 90% of the time, $P(flag|spam) = 90\%$
- It mistakenly flags normal emails 5% of the time: $P(flag|not\ spam) = 5\%$

An email is flagged as spam. What is the probability that the email is actually spam, i.e. $P(spam|flag)$?

Example 7: Email Spam Filter

An email system automatically classifies emails as "Spam".

- 30% of incoming emails are actually spam.
- The filter correctly detects spam 90% of the time, $P(flag|spam) = 90\%$
- It mistakenly flags normal emails 5% of the time: $P(flag|not\ spam) = 5\%$

An email is flagged as spam. What is the probability that the email is actually spam, i.e. $P(spam|flag)$?

$$\begin{aligned} P(spam|flag) &= \frac{P(flag|spam)P(spam)}{P(flag|spam)P(spam) + P(flag|not\ spam)P(not\ spam)} \\ &= \frac{90\% \cdot 30\%}{90\% \cdot 30\% + 5\% \cdot 70\%} = 89\% \end{aligned}$$

Example 8: Winning a Special Prize

In a school lottery:

- 10% of students buy a ticket.
- Only students who buy a ticket can win.
- Among students who buy a ticket, 5% win a prize.

Let:

- $P(ticket)$ the probability of buying a ticket.
- $P(win)$ the probability of winning.

A randomly selected student is found to have won a prize. What is the probability that this student bought a ticket? $P(ticket|win)$?

Example 8: Winning a Special Prize

In a school lottery:

- 10% of students buy a ticket.
- Only students who buy a ticket can win.
- Among students who buy a ticket, 5% win a prize.

Let:

- $P(ticket)$ the probability of buying a ticket.
- $P(win)$ the probability of winning.

A randomly selected student is found to have won a prize. What is the probability that this student bought a ticket? $P(ticket|win)$?

$$
\begin{aligned}
P(ticket|win) &= \frac{P(win|ticket)P(ticket)}{P(win)} \\
&= \frac{P(win|ticket)P(ticket)}{P(win|ticket)P(ticket) + P(win|non\ ticket)P(non\ ticket)} \\
&= \frac{5\% \cdot 10\%}{5\% \cdot 10\% + 0\% \cdot 90\%} = 100\%
\end{aligned}
$$

Example 9: Laboratory Chemical Test

In a chemistry lab:

- 8% of samples contain a rare substance.
- The test detects the substance 98% of the time:
  $P(Positive|Substance) = 0.98$
- The test gives a false positive only 0.2% of the time:
  $P(Positive|No\ Substance) = 0.002$

A randomly chosen sample tests positive. What is the probability that the sample actually contains the substance? $P(Substance|Positive)$

Example 9: Laboratory Chemical Test

In a chemistry lab:

- 8% of samples contain a rare substance.
- The test detects the substance 98% of the time: $P(Positive|Substance) = 0.98$
- The test gives a false positive only 0.2% of the time: $P(Positive|No\ Substance) = 0.002$

A randomly chosen sample tests positive. What is the probability that the sample actually contains the substance? $P(Substance|Positive)$

$$\begin{aligned} P(Substance|Positive) &= \frac{P(Positive|Substance)P(Substance)}{P(Positive)} \\ &= \frac{P(Positive|Substance)P(Substance)}{P(Positive|Sub)P(Sub) + P(Positive|non\ Sub)P(non\ Sub)} \\ &= \frac{98\% \cdot 8\%}{98\% \cdot 8\% + 0.2\% \cdot 92\%} = 97.7\% \end{aligned}$$

Example 10: Airport Security Alert

At a large airport:

- Only 0.5% of passengers are carrying a prohibited item, $P(R) = 0.005$
- The scanner detects prohibited items 99% of the time: $P(alarm|R) = 0.99$
- The scanner falsely alarms 3% of the time for innocent passengers: $P(alarm|not\ R) = 0.03$

A passenger triggers the alarm. What is the probability the passenger is actually carrying a prohibited item? $P(R|alarm)$

Example 10: Airport Security Alert

At a large airport:

- Only 0.5% of passengers are carrying a prohibited item, $P(R) = 0.005$
- The scanner detects prohibited items 99% of the time: $P(alarm|R) = 0.99$
- The scanner falsely alarms 3% of the time for innocent passengers: $P(alarm|not\ R) = 0.03$

A passenger triggers the alarm. What is the probability the passenger is actually carrying a prohibited item? $P(R|alarm)$

$$\begin{aligned} P(R|alarm) &= \frac{P(alarm|R)P(R)}{P(alarm)} \\ &= \frac{P(alarm|R)P(R)}{P(alarm|R)P(R) + P(alarm|non\ R)P(non\ R)} \\ &= \frac{99\% \cdot 0.5\%}{99\% \cdot 0.5\% + 3\% \cdot 99.5\%} = 14\% \end{aligned}$$

Even with 99% detection accuracy, the probability that an alarm means real danger is only 14%. Most alarms are false.

Example 11

In a casino in Blackpool there are two slot machines: one that pays out 10 % of the time, and one that pays out 20 % of the time. Obviously, you would like to play on the machine that pays out 20 % of the time but you do not know which of the two machines is the more generous. You thus adopt the following strategy: you assume initially that the two machines are equally likely to be the generous machine. You then select one of the two machines at random and put a coin into it. Given that you loose that first bet estimate the probability that the machine you selected is the more generous of the two machines.

Example 11

In a casino in Blackpool there are two slot machines: one that pays out 10 % of the time, and one that pays out 20 % of the time. Obviously, you would like to play on the machine that pays out 20 % of the time but you do not know which of the two machines is the more generous. You thus adopt the following strategy: you assume initially that the two machines are equally likely to be the generous machine. You then select one of the two machines at random and put a coin into it. Given that you loose that first bet estimate the probability that the machine you selected is the more generous of the two machines.

- G = the machine you chose is the generous one (20% payout)
- L = the machine you chose is the less generous one (10% payout)
- B = you lose the first bet

We know from the text that P(G)=P(L)=0.5 (machine randomly chosen).

$$\begin{aligned} P(G|B) &= \frac{P(B|G)P(G)}{P(B)} \\ &= \frac{P(B|G)P(G)}{P(B|G)P(G) + P(B|L)P(L)} \\ &= \frac{0.8 \cdot 0.5}{0.8 \cdot 0.5 + 0.9 \cdot 0.5} = 47\% \end{aligned}$$