



Probability Axioms, Conditional Probability and Bayes

Statistics I · Lecture 5

Today

- Axioms and corollaries
- Conditional probability and independence
- Law of Total Probability and Bayes theorem
- Practice, from  to 

Probability Axioms

Probability

Andrey Kolmogorov defined a set of characteristics that any probability P measure should have, these are called the Kolmogorov's axioms (1933):

1. $P(A) \in \mathbb{R}$ and $P(A) \geq 0$, for any event $A \subseteq \Omega$.
2. $P(\Omega) = 1$
3. For any A and B disjoint, $P(A) + P(B) = P(A \cup B)$

Corollary

Let A and B be some events in Ω

1. $P(\emptyset) = 0$
2. $P(A^c) = 1 - P(A)$
3. $A \subset B \Rightarrow P(A) \leq P(B)$
4. $0 \leq P(A) \leq 1, \forall A$
5. $P(A \setminus B) = P(A) - P(A \cap B)$

Corollary

$$6. P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) +$$

$$7. \sum_{i \neq j \neq k} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right)$$

Conditional Probability

Conditional probability, as the wording implies, means the probability of something happening **given** something else has happened. Now, note “something” here makes reference to an *event*.

$$P(A|B)$$

It reads *the probability of A, given B*.

Conditional Probability

Note that if we think on sets, saying *given* B we are immediately excluding everything that could have happened if B did not happen, and therefore our Universal set is no longer Ω , but B .

What we are looking for are, among the events that live in B , how many of those live in A (because those would *trigger* event A). Actually, we are interested on the relative measure of those outcomes, compared to the whole size of B :

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

Example

Consider a factory that makes 10 wrenches . Among those, we know that 2 have imperfections. Suppose you intend to remove, randomly, 2  from the lot (of 10). Consider the following events:

$A = \{\text{The first :wrench: is faulty}\}$ $B = \{\text{The second :wrench: is faulty}\}$

What if we want to compute $P(B)$? For a correct assessment for B , we would better have some information on the realization of A !

Example

If the first  was faulty, then A happened. If the first  was *ok*, then A^c happened, and therefore we can compute $P(B|A)$ and $P(B|A^c)$. We are assuming that we are removing these  without replacing them.

Let's see why this last detail (replacing the ) is so relevant before going on.

Example: With replacement

Initial set:

1	2	3	4	5	6	7	8	9	10
									

Remove one (if randomly you do not know which), but after you remove you can see what happened, let's take out 7.

1	2	3	4	5	6	7	8	9	10
									

We observe, and *voilà* it was a fine wrench . If we replace it though, we would be picking from

1	2	3	4	5	6	7	8	9	10
									

That is in the exact same conditions we made our first choice, and therefore what happens with the first pick is irrelevant: These events are now independent!

Example: with replacement

- With A , we know that when we pick the second wrench (B), in the box there are 2  and 8 .
- With A^c , we know that when we pick the second wrench (B), in the box there are 2  and 8 .

The probability of getting a  is the same in each scenario!

$$P(B|A) = P(B|A^c)!$$

$$P(B|A) = \frac{2}{10} = 0.2 \text{ and } P(B|A^c) = \frac{2}{10} = 0.2$$

Example: no replacement

Initial set:

1	2	3	4	5	6	7	8	9	10
									

Remove one (if randomly you do not know which is broken)

1	2	3	4	5	6	7	8	9	10
									

We observe, and *voilà* it was a broken wrench  A happened!

1	2	3	4	5	6	7	8	9	10
									

We observe, and *voilà* it was a fine wrench  A^c happened!

Example: no replacement

- With A , we know that when we pick the second wrench (B), in the box there are 1 💣 and 9 🔧.
- With A^c , we know that when we pick the second wrench (B), in the box there are 2 💣 and 9 🔧.

The probability of getting a 💣 is different in each scenario!

$$P(B|A) \neq P(B|A^c)!$$

$$P(B|A) = \frac{1}{9} = 0.111 \text{ and } P(B|A^c) = \frac{2}{9} = 0.222$$

Conclusion Conditional Probability

So formally

Definition

Let $A, B \subset \Omega$, then we say the probability of A given B is the conditional probability:

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

Note that from here, we can obtain also $P(A \cap B) = P(A|B)P(B)$. In both situations we need $P(B) \neq 0$.

Corollary

It follows that the identity

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

or

$$P(A \cap B) = P(B|A)P(A)$$

with $P(A) \neq 0$ also holds true.

Corollary

Now let's think on $P(A \cap B \cap C)$:

- $P(A \cap B \cap C) = P(A \cap (B \cap C))$
- $P(A \cap B \cap C) = P(A|B \cap C)P(B \cap C)$
- $P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$

Corollary

Note that given the commutativity of the intersection, we could have obtained also:

- $P(A \cap B \cap C) = P(A|B \cap C)P(C|B)P(B)$
- $P(A \cap B \cap C) = P(B|A \cap C)P(A|C)P(C)$
- $P(A \cap B \cap C) = P(B|A \cap C)P(C|A)P(A)$
- $P(A \cap B \cap C) = P(C|A \cap B)P(A|B)P(B)$
- $P(A \cap B \cap C) = P(C|A \cap B)P(B|A)P(A)$

And to make sense of all of this we need $P(X) > 0$, $P(X \cap Y) > 0$ with $X, Y \in \{A, B, C\}$.

Example

Consider a region with 1,000 adults. Their job data is captured by the following table:

	Employed	Unemployed	Total
Women	470	55	525
Men	430	45	475
Total	900	100	1,000

1. Randomly selecting a person in this region, what is the probability this person is:
 1. Woman
 2. Unemployed
 3. Unemployed woman

Example

Let's define the events:

$$W = \{Woman\}, M = \{Man\}, U = \{Unemployed\}$$

1. Woman: $P(W) = \frac{525}{1000} = 0.525$

2. Unemployed: $P(U) = \frac{100}{1000} = 0.1$

3. Unemployed woman: $P(W \cap U) = \frac{55}{1000} = 0.055$

Example

2. A citizen is randomly chosen from the population, and it happens to be a woman. What is the probability she is unemployed?

$$P(U|W) = \frac{P(U \cap W)}{P(W)} = \frac{0.055}{0.525} = 0.105$$

3. A citizen is randomly chosen from the population, and it happens to be unemployed. What is the probability this citizen is a woman?

$$P(W|U) = \frac{P(W \cap U)}{P(U)} = \frac{0.055}{0.1} = 0.55$$

Independent Events

Definition

Two events A and $B \subset \Omega$, are probabilistically independent if and only if:

$$P(A \cap B) = P(A)P(B)$$

Independent Events

From the definition of independence, we can obtain several properties.

Let A and B independent events with $P(A)P(B) > 0$:

1. $P(A|B) = P(A)$ and $P(B|A) = P(B)$ (remember conditional example with replacement)
2. A^c and B are independent, as well as A and B^c , and even A^c and B^c .
3. If A and B are incompatible, they cannot be independent.
$$P(A \cap B) = P(\emptyset) = 0 \neq P(A)P(B)$$
4. Any event is independent of Ω and $\emptyset$.

Example

Are W and U from the previous example independent?

$$P(W) = 0.525, P(U) = 0.1, P(W|U) = 0.55, P(U|W) = 0.105.$$

Note that $P(W) \neq P(W|U)$ and $P(U) \neq P(U|W)$. Therefore, **they cannot be independent.**

Example

Consider a die  that is thrown twice. Consider the following two events:

$A = \{\text{The die shows an odd number the first time}\}$

$B = \{\text{The die shows a number } > 4 \text{ the second time}\}$

Are A and B independent events?

Example

In this case, $\Omega = \{(x, y) \in \mathbb{N}^2 \mid x, y \leq 6\}$ with $\#\Omega = 6^2 = 36$

- $P(A) = \frac{18}{36} = \frac{1}{2}$
- $P(B) = \frac{12}{36} = \frac{1}{3}$
- $P(A \cap B) = \frac{1}{6} = P(A)P(B)$
- $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{2} = P(A)$
- $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{1}{3} = P(B)$
- They are independent events!

Remark

Two events being independent is not the same that they being incompatible:

<i>A</i> and <i>B</i> independent	<i>A</i> and <i>B</i> incompatible
$P(A \cap B) = P(A)P(B)$	$P(A \cap B) = 0$
$P(A B) = P(A)$ and $P(B A) = P(B)$	$P(A B) = 0$ and $P(B A) = 0$

Example

Let A and B be two events such that: $P(A) = 0.6$, $P(B) = t$, and $P(A \cup B) = 0.8$

Find t such that A and B are:

1. Mutually exclusive or incompatible.
2. Independent.

Example

1. In this case, what we need is that $P(A \cap B) = 0$.

From probability theory, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

and therefore we get:

$$0.8 = 0.6 + t \Rightarrow t = 0.2$$

Example

2. To make A and B independent, we need that

$$P(A \cap B) = P(A)P(B) = 0.6t:$$

Using the same identity we just used:

$$0.8 = 0.6 + t - 0.6t \Rightarrow t = 0.5$$

Law of Total Probability

Theorem

Let $\{A_i\}_{i=1}^n$ be a partition of Ω , so $\{A_i\} \subseteq \mathcal{P}(\Omega)$. Then, for any $B \subset \Omega$, it holds that:

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

Example

Consider a financial institution that sells two products, α and β , with very high yields. It is known that, among its clients, 10% invest a share of their wealth in α and the rest in β . From those who invest in α , 70% manage to get returns above the market. From among those who do not invest in α , 55% get returns above the market. Randomly choosing a client of this firm, find the probability this customer gets a return above the market.

Example

Let's define the events:

- A_1 the client invest in α
- A_2 the client invest in β
- B has returns above the market.

Matching with the available data we obtain:

$$P(A_1) = 0.1, P(A_2) = 0.9, P(B|A_1) = 0.7, \text{ and } P(B|A_2) = 0.55.$$

From the Law of Total Probability:

Example

$$P(B) = \sum_{i=1}^2 P(A_i \cap B)$$

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2)$$

$$P(B) = 0.7 \times 0.1 + 0.55 \times 0.9 = 0.565$$

Bayes Theorem

Bayes Theorem

Let events $A_1, A_2, \dots, A_n$ with $n \in \mathbb{N}$ a partition of Ω , then, for any event $B \subset \Omega$, with $P(B) > 0$:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

with $i = 1, 2, \dots, n$

Note that this is a consequence of the Law of Total Probability.

Bayes Theorem

On the other side, $\sum_i P(A_i) = 1$ and $\sum_i P(A_i|B) = 1$

Bayes Theorem has been widely used in economics, in biomedical sciences, and social sciences when looking for causality.

If event B represents *consequences* and event A_i probable *cause*, Bayes Theorem allows to assess the probability of this cause ($P(A_i)$).

Example

Let's go back to the previous example, about our investors.

Let's compute the probability that the client invested his money on product β , but given that the client had returns above the market (event B).

Example

If the customer invested in β , then the event we are trying to is A_2 , but conditional on event B , $P(A_2|B)$:

$$P(A_2|B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{P(A_2) \times P(B|A_2)}{\sum_i P(A_i) \times P(B|A_i)}$$

We knew from the previous exercise that $P(B) = 0.565$, and therefore we obtain:

$$P(A_2|B) = \frac{0.9 \times 0.55}{0.565} = 0.876$$

Example

How do we interpret this?

The probability that the client invested in β , given that he had a return above the market, is 0.876.

Example

All these computations can be very easy with the help of the following table:

A_i	$P(A_i)$	$P(B A_i)$	$P(A_i)P(B A_i)$	$P(A_i B)$
A_1	0.1	0.7	0.07	0.124
A_2	0.9	0.55	0.495	0.876
	1		0.565	1

A test that is right 99% of the time

A screening test for a rare condition. Move the sliders and read off the only number a patient cares about:
given a positive test, what is the chance of actually being ill?



Prevalence **1.0%**



Sensitivity **99.0%** $P(+ | \text{ill})$



Specificity **99.0%** $P(- | \text{healthy})$



$P(\text{ill} | \text{positive}) = 50.0\%$

In 100 000 people:

- truly ill and positive: 990
- healthy but positive: 990
- ill but missed: 10

What the chart showed

At 99% sensitivity **and** 99% specificity, a positive result on a condition affecting 1 in 100 people leaves you only about **50%** likely to be ill. Push prevalence down to 1 in 1000 and the answer collapses to roughly **9%**.

The reason is in the two blocks of the bar: the healthy group is so much larger that its **1% error rate still produces more positives** than the ill group produces in total.

$P(\text{ill} \mid +)$ is not $P(+ \mid \text{ill})$. Swapping them is the single most common probability mistake outside this classroom, and Bayes is what keeps them apart.

Example

Let's verify now if the event A_1 and B^c are independent or not!

According to the definition of independence: $P(A_1 \cap B^c) = P(A_1)P(B^c)$

- $P(A_1 \cap B^c) = P(A_1) \times P(B^c|A_1) = 0.1 \times 0.3 = 0.03$
- $P(A_1) \times P(B^c) = 0.1 \times (1 - 0.565) = 0.0435$
- Then:

$$P(A_1 \cap B^c) = P(A_1) \times P(B^c) \Leftrightarrow 0.03 \neq 0.0435$$

Then A_1 and B^c are not independent.

? Axioms, Independence and Bayes ·

Question 1

A and B are incompatible, with $P(A) > 0$ and $P(B) > 0$. Then they are:

A. independent

B. independent only if disjoint

C. never independent

D. impossible to classify

✓ **C.** Incompatible means $P(A \cap B) = 0$, but independence needs $P(A \cap B) = P(A)P(B) > 0$. The two cannot hold together.

? Axioms, Independence and Bayes ·

Question 2

$P(A) = 0.5$, $P(B) = 0.4$ and $P(A \cap B) = 0.2$. Then A and B are:

- A. incompatible
- B. neither
- C. independent
- D. both

✓ **C.** $P(A)P(B) = 0.5 \times 0.4 = 0.2 = P(A \cap B)$ ✓



Axioms, Independence and Bayes ·

Question 3

A disease affects 2% of the population. A test is positive for 95% of ill people, and also for 5% of healthy people.

A randomly chosen person tests positive. What is the probability that this person is ill?

Axioms, Independence and Bayes · Solution

Let D be “ill” and $+$ be “tests positive”. By the Law of Total Probability:

$$P(+) = 0.95 \times 0.02 + 0.05 \times 0.98 = 0.019 + 0.049 = 0.068$$

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{0.019}{0.068} \approx 0.279$$

Only about **28%**, even though the test is right 95% of the time. The disease is rare, so most positives come from the large healthy group.

Practice

Practice

The exercises go from easy to hard:

-  warm-up: one formula, one step
-  standard: two or three steps
-  exam level: this is what you will find in the midterm
-  stretch: you need to combine several things

Practice · Exercise 1

Let A and B be two events with $0 < P(A) < 1$ and $0 < P(B) < 1$. You know that $A \subset B$.

What is $P[(A \cup B) \cap B^c]$?

A. 0

B. $P(A)$

C. $P(B)$

D. 1

☒ **A.** If $A \subset B$ then $A \cup B = B$, and $B \cap B^c = \emptyset$. Draw the Venn diagram if you doubt it.

Practice · Exercise 2

Let A and B be two events such that $P(A) = 0.75$, $P(B) = 0.5$ and $P(A \cup B) = 1$.

Find $P(A|B)$.

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.75 + 0.5 - 1 = 0.25$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.25}{0.5} = 0.5$$

Practice · Exercise 3

Let $P(A) = a$ and $P(B) = b$, with $0 < a < 1$ and $0 < b < 1$. True or false, the probability that neither A nor B happen is:

(a) $1 - a - b + ab$, if A and B are independent.

(b) 1, if A and B are complementary.

(a)  True.

$P(A^c \cap B^c) = P(A^c)P(B^c) = (1 - a)(1 - b) = 1 - a - b + ab$, since the complements of independent events are also independent.

(b)  False. If they are complementary, one of them always happens, so the probability that none happens is 0.

Practice · Exercise 4

Let A, B, C be events such that:

$A \cup B \cup C = \Omega$, $P(A) = 0.3$, $P(C) = 0.5$, $P(B^c) = 0.7$, $A \cap B = \emptyset$ and $B \cap C = \emptyset$.

Find $P(A \cap C)$.

$P(B) = 1 - 0.7 = 0.3$. B does not intersect A nor C , so all the intersections with B vanish:

$$1 = P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap C)$$

$$1 = 0.3 + 0.3 + 0.5 - P(A \cap C) \Rightarrow P(A \cap C) = 0.1$$

Practice · Exercise 5

A and B are independent, and A has a probability twice as large as B . The probability that at least one of them happens is 0.5.

Find $P(B)$.

Let $p = P(B)$, so $P(A) = 2p$ and, by independence, $P(A \cap B) = 2p^2$:

$$0.5 = 2p + p - 2p^2 \Leftrightarrow 2p^2 - 3p + 0.5 = 0 \Leftrightarrow p = \frac{3 \pm \sqrt{5}}{4}$$

$\frac{3+\sqrt{5}}{4} \approx 1.31$ is not a probability! So $P(B) = \frac{3-\sqrt{5}}{4} \approx 0.191$ and $P(A) \approx 0.382$.

Practice · Exercise 6

5% of the students are excluded from continuous assessment , and of these, 98% end up with a final grade above 13. Of those not excluded, 10% end up with a final grade below 13.

Pick a student randomly at the end of the semester.

(a) What is the probability of a grade above 13, given the student was in continuous assessment?

(b) Knowing the student got a grade above 13, what is the probability that the student was excluded?



Practice · Exercise 6 · Solution

Let E be “excluded” and G “grade above 13”. $P(E) = 0.05$,
 $P(G|E) = 0.98$, $P(G^c|E^c) = 0.1$.

(a) $P(G|E^c) = 1 - 0.1 = 0.90$

(b) By the Law of Total Probability,

$$P(G) = 0.05 \times 0.98 + 0.95 \times 0.90 = 0.049 + 0.855 = 0.904$$

$$P(E|G) = \frac{P(G|E)P(E)}{P(G)} = \frac{0.049}{0.904} \approx 0.0542$$



Homework

Problem set **2.1**, Questions 3 to 10. Questions 8 to 10 are Bayes, like Exercise 6, and we will come back to them before the midterm.

Bibliography

- Murteira, B.; Ribeiro C.; Silva, J. and Pimenta, C. (2010) Introdução à Estatística (2ª Edição). McGraw-hill.
- Paulino C.D.; Branco J.A. (2005). Exercícios de Probabilidade e Estatística. Escolar Editora.
- Pedrosa, A.; Gama, S. (2004). Introdução Computacional à Probabilidade e Estatística. Porto Editora.