Foundations of Machine Learning: Week 1: Conditional Probability

Professor Christopher Yau
christopher.yau@manchester.ac.uk

October 21, 2020

Foundations of Machine Learning:Week 1: Conditional Probability

Foundations of Machine Learning:
Week 1: Conditional Probability
Professor Christopher Yau
christopher.yau@manchester.ac.uk
October 21, 2020

Conditional Probability

Conditional probabilities are used to describe the probability of event occurring given that another event has already occurred.

Examples:

- What is the probability that a football team will win a match *given* that they score first,
- What is the probability of a vehicle accident *given* that there is snow.

For events E and F, we use the notation $P(E|F)$ to denote the *conditional probability* of E given F.

Conditional Probability

Conditional probabilities are used to describe the probability of event occurring given that another event has already occurred.

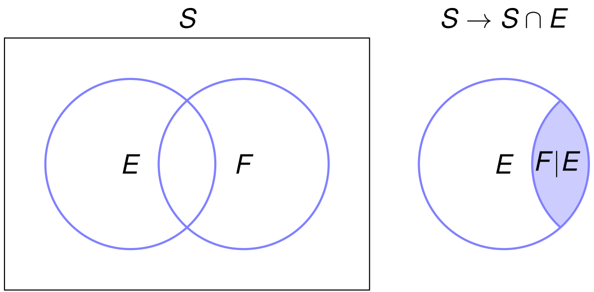
Examples:

- What is the probability that a football team will win a match *given* that they score first,
- What is the probability of a vehicle accident *given* that there is snow.

For events E and F, we use the notation $P(E|F)$ to denote the *conditional probability* of E given F.

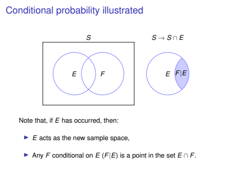
In this lecture, we will look at *conditional probabilities*.
This is the probability of an event occurring given another event has already occurred.
For example:
What is the probability that a football team will win a match *given* that they score first.
What is the probability of a vehicle accident *given* that there is snow.
For events E and F, we use the notation $P(E|F)$ to denote the *conditional probability* of E given F.

Conditional probability illustrated



Note that, if E has occurred, then:

- E acts as the new sample space,
- Any F conditional on E ($F|E$) is a point in the set $E \cap F$.



What does this mean graphically?
Using our Venn diagram, when we condition on an event occurring, say E, we can think of the sample space collapsing from S to E.
The event E is like our new sample space because things not in E can no longer occur given E has happened.
Therefore, any future event F must overlap E and must be a point in the set $E \cap F$.

Chain Rule

The **chain rule** (or the product rule) tells us that:

$$P(E \overbrace{\cap}^{\text{AND}} F) = P(E|F)P(F).$$

This tells us that the probability of two events occurring is the probability of one event occurring times the probability of the second event given the first has occurred.

Chain Rule
The **chain rule** (or the product rule) tells us that:
$$P(E \overbrace{\cap}^{\text{AND}} F) = P(E|F)P(F).$$
This tells us that the probability of two events occurring is the probability of one event occurring times the probability of the second event given the first has occurred.

Conditional probabilities can be combined using the *chain rule* (sometimes known as the product or multiplication rule).
It tells us that the probability that two events E and F occurring is the probability that one event occurs, say F, and then E happens given F has happened.

Chain Rule

Note, through symmetry,

$$P(F|E)P(E) = P(E \cap F) = P(E|F)P(F).$$

Note that standard axioms for probability extend to conditional probability, e.g.

$$P(\overline{E}|F) = 1 - P(E|F).$$

Chain Rule

Note, through symmetry,

$$P(F|E)P(E) = P(E \cap F) = P(E|F)P(F).$$

Note that standard axioms for probability extend to conditional probability, e.g.

$$P(\overline{E}|F) = 1 - P(E|F).$$

Note that by symmetry, the complement is also true, the probability that two events E and F occurs is the probability that E first, and then F happens given E has happened.

Note that the addition and complement laws also apply to conditional probabilities as well so the probability that E does not occur given F is given by one minus the probability that E occurs given F.

Independence

If for two events E and F

$$P(E|F) = P(E)$$

then E is said to be **independent** of F.

That is, the fact that event F occurs does not affect whether event E happens.

This implies:

$$P(E \cap F) = P(E)P(F).$$

Independence

If for two events E and F

$$P(E|F) = P(E)$$

then E is said to be **independent** of F.

That is, the fact that event F occurs does not affect whether event E happens.

This implies:

$$P(E \cap F) = P(E)P(F).$$

When discussing conditional probabilities, it is important to note the concept of *independence*.
When an event is said to be independent of another, it means that the fact that one event has occurred has no impact of the occurrence of the other.
This means that the conditional probability of E given F is simply the probability of E if E is independent of F.
When two events are independent, the probability of both events occurring is simply given by the product of their individual probabilities.
Noting independence is important in real-world modelling as it helps to simplify computation. One does not need to be concerned about the historical dependence effects if an event is independent of the others.

Probability Tables

A **probability table** can be a convenient way to enumerate the probabilities of pairs of events.

	A	$\bar{A}$	
B	$P(A \cap B)$	$P(\bar{A} \cap B)$	$P(B)$
$\bar{B}$	$P(A \cap \bar{B})$	$P(\bar{A} \cap \bar{B})$	$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	1

Probability Tables

A **probability table** can be a convenient way to enumerate the probabilities of pairs of events.

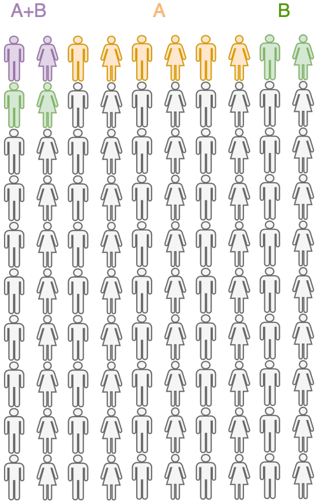
	A	$\bar{A}$	
B	$P(A \cap B)$	$P(\bar{A} \cap B)$	$P(B)$
$\bar{B}$	$P(A \cap \bar{B})$	$P(\bar{A} \cap \bar{B})$	$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	1

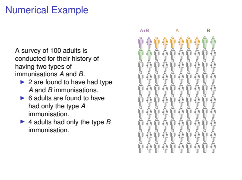
For hand calculations with two events, it can be useful to draw a probability table.
Each row and column corresponds to one of the events occurring or not, and each cell denotes the combination of those events.
The sum of the four cells must add up to one.
The totals of each row or column add up to give the probability of each event in isolation. This is known as the marginal probability which we will discuss in the next lecture.

Numerical Example

A survey of 100 adults is conducted for their history of having two types of immunisations A and B.

- 2 are found to have had type A and B immunisations.
- 6 adults are found to have had only the type A immunisation.
- 4 adults had only the type B immunisation.





Lets consider this idea of conditional probability with a numerical example.
A survey of 100 adults is conducted for their history of having two types of immunisations A and B.

- 2 are found to have had type A and B immunisations.
- 6 adults are found to have had only the type A immunisation.
- 4 adults had only the type B immunisation.

Solution

What is the probability that an adult has had *at least* one immunisation?

Construct a table of counts:

	A	$\overline{A}$	
B	2	4	(6)
$\overline{B}$	6	(88)	(94)
	(8)	(100-8)	100

We want $P(A \cup B)$ and we could calculate this as:

$$P(A \cup B) = P(\overline{A} \cap B) + P(A \cap \overline{B}) + P(A \cap B)$$

which gives $P(A \cup B) = 4/100 + 6/100 + 2/100 = 12/100 = 3/25$.

Via the addition law:

$$P(A \cup B) = P(A) + P(B) + P(A \cap B) = 8/100 + 6/100 - 2/100 = 12/100$$

Solution
What is the probability that an adult has had at least one immunisation?
Construct a table of counts:

	A	$\overline{A}$	
B	2	4	(6)
$\overline{B}$	6	(88)	(94)
	(8)	(100-8)	100

We want $P(A \cup B)$ and we could calculate this as:
$$P(A \cup B) = P(\overline{A} \cap B) + P(A \cap \overline{B}) + P(A \cap B)$$
which gives $P(A \cup B) = 4/100 + 6/100 + 2/100 = 12/100 = 3/25$.
Via the addition law:
$$P(A \cup B) = P(A) + P(B) + P(A \cap B) = 8/100 + 6/100 - 2/100 = 12/100$$

We are asked to compute the probability that an adult has had at least one immunisation.
First we can construct a table of counts of the number of people in each category. We can convert these into probabilities by dividing by the number of people (100).
We would like to compute the probability that someone has immunisation A or B.
We can compute this in a number of ways:

- We can add up the three cells which correspond to our events of interest.
- We could use the addition law.

Solution

What is the probability that an adult has had immunisation A given that they have had immunisation B?

We want $P(A|B)$ and we can calculate this using the law of conditional probability:

$$P(A|B) = P(A \cap B)/P(B) = (2/100)/(6/100) = 2/6 = 1/3.$$

Note - rearranged from $P(A \cap B) = P(A|B)P(B)$

Solution

What is the probability that an adult has had immunisation A given that they have had immunisation B?

We want $P(A|B)$ and we can calculate this using the law of conditional probability:

$P(A|B) = P(A \cap B)/P(B) = (2/100)/(6/100) = 2/6 = 1/3.$

Note - rearranged from $P(A \cap B) = P(A|B)P(B)$

Now suppose we wish to find the probability that someone has the immunisation A given they have had immunisation B.
This is a conditional probability question and we want the probability of A given B.
From the definition of conditional probability we know the probability of A and B is equal to the conditional probability of A given B times the probability of B.
We can re-arrange this to get the expression for the conditional probability in terms of the others.
So the probability that someone has a type A immunisation, given they have had type B is given by the probability that they have both divided by the probability that they had B irrespective of whether they had A.

Numerical Example

A family has two children.

What is the probability that both are boys given that at least one is a boy?

Numerical Example
A family has two children.
What is the probability that both are boys given that at least one is a boy?

A family has two children.
We are asked to compute the probability that both are boys given that at least one is a boy?

Solution

The two possible outcomes are B (Boy) and G (Girl)

$$S = \{GG, GB, BG, BB\}$$

The question requires to compute $P(BB|\text{at least one B})$

Assuming equal probabilities, i.e. 1/4, we have:

$$P(GG) = P(BG) = P(GB) = P(BB) = 1/4$$

$$P(BB|\{BG \cup GB \cup BB\}) = \frac{P(BB \cap (BG \cup GB \cup BB))}{P(BG \cup GB \cup BB)} = \frac{1/4}{3/4} = 1/3$$

Note - $P(BB \cap (BG \cup GB \cup BB)) = P(BB) = 1/4$

Solution
The two possible outcomes are B (Boy) and G (Girl)
$S = \{GG, GB, BG, BB\}$
The question requires to compute $P(BB|\text{at least one B})$
Assuming equal probabilities, i.e. 1/4, we have:
$P(GG) = P(BG) = P(GB) = P(BB) = 1/4$
$P(BB|\{BG \cup GB \cup BB\}) = \frac{P(BB \cap (BG \cup GB \cup BB))}{P(BG \cup GB \cup BB)} = \frac{1/4}{3/4} = 1/3$
Note - $P(BB \cap (BG \cup GB \cup BB)) = P(BB) = 1/4$

We first need to enumerate the appropriate sample space. In this case, the two children can either be two boys, two girls or a boy and a girl in any order.
The question requires us to compute the probability of two boys given there is at least one boy in the family. Lets assume each of these four possibilities is equally likely so they have 1 in 4 chance of occurring.
The event of at least one boy is the subset of outcomes: boy-girl, girl-boy and boy-boy.
We can therefore rewrite the conditional probability of two boys given at least one boy using the chain rule as the joint probability of two boys and at least one boy divided by the probability of at least one boy.
The joint probability is 1/4 since the only overlapping event is two-boys which has probability 1/4. The denominator term is 3/4 since there are three events out of four.
Therefore the conditional probability is 1/3 which makes sense since, given that BG, GB and BB happen, the probability of BB occurs is 1 out of those 3 events.

Recap: Conditional probability

The joint probability that both events E **and** F occur is:

$$P(E \cap F) = P(F|E)P(E) = P(E|F)P(F)$$

where $P(F|E)$ and $P(E|F)$ are **conditional** probabilities.

Note that, if E has occurred, then

- $F|E$ is a point in the set $E \cap F$,
- E is the new sample space.

Recap: Conditional probability

The joint probability that both events E **and** F occur is:

$P(E \cap F) = P(F|E)P(E) = P(E|F)P(F)$

where $P(F|E)$ and $P(E|F)$ are **conditional** probabilities.

Note that, if E has occurred, then

- $F|E$ is a point in the set $E \cap F$,
- E is the new sample space.

So to summarise, the joint probability that two events E and F occur is given by the probability that one event occurs, say E and then the conditional probability of the second event F happening given E has happened.

Conditional probabilities will be very important for when we later consider Bayes' Theorem.

END LECTURE

END LECTURE