Basic probability theory

Chrysafis Vogiatzis

Lecture 3

Learning objectives

After this lecture, we will be able to:

- Recall and explain the basic properties that probability has.
- Calculate the probability of an event.
- Apply set operations in probability calculations.
- Define and provide examples of conditional probabilities.
- Apply the conditional probability formula.
- Recognize independence.

Motivation: Will I miss my flight?

Flying from Urbana-Champaign almost always requires a layover in another airport. For example, flying to New York City usually is done through Chicago with two legs: Urbana-Champaign to Chicago, and Chicago to New York City. My layover is only 45 minutes in Chicago, so I am naturally worried about making my connection. I would feel much better if I knew whether my first flight leaves on time or not. What is the probability that I make my second flight given that my first flight is delayed by 15 minutes or more?

Motivation: Data collection

A company has undertaken the large effort of contact tracing and testing for COVID-19 in the Urbana-Champaign area. It is expected that from the people that leave in the area, 1% has been in close contact with an already known case of COVID-19, 15% has been working in close contact with multiple people as an essential worker, and 6% has traveled to a location (in and outside the USA) with high risk of contagion. A random person is selected for a test, but will only be administered the test if they fall within one of the three categories above. What is the probability that the person will get the test?

Probabilities

Definition

As a continuation from the motivation in the previous lecture, there are two interpretations of probabilities:

1. relative frequency of favorable outcomes versus all outcomes.
2. subjective "degree of belief".

In both cases, we can use *basic probability theory* to calculate the likelihood of an event happening or not. Once again, recall the definition of probability:

Definition 1 (Probability) *With every event, we associate a real number called probability to represent the likelihood of that event happening. Probabilities satisfy three main rules* [1]*:*

1. *$P(E) \geq 0$, for any event E.*
2. *If an event E comprises the whole sample space (in which case, we write that $E = S$), then $P(E) = 1$.*
3. *If $E_1, E_2, \ldots, E_m$ are m mutually exclusive events, then*

$$P\left(E_1 \cup E_2 \cup \ldots \cup E_m\right) = P\left(E_1\right) + P\left(E_2\right) + \ldots + P\left(E_m\right),$$

or even more concisely:

$$P\left(\bigcup_{i=1}^{m} E_i\right) = \sum_{i=1}^{m} P(E_i).$$

[1] Also known as the Kolmogorov axioms of probability.

The first two axioms imply that probability is a real number in $[0, 1]$ [2], where 0 is an *impossible event* (one that can never happen) and 1 signals a *certain event* (one that will always happen) [3].

[2] We sometimes present probability as a percentage (%): for example, a probability of 0.4 can be also written as a probability of 40%.

[3] Consider a sample space S that consists of three events: A, B, C. Then the event that neither A nor B nor C happen is *impossible*; the event that A or B or C happens is *certain*.

Probabilities of mutually exclusive events

Assume that two events E_1, E_2 are mutually exclusive: for example, let E_1 be the event that you get an A in IE 300, and E_2 the probability that you get an $A-$. Your *personal belief* is that you have a 30% "chance" at an A and a 20% "chance" at an $A-$. Then, the probability that you get *at least an $A-$ in the class* is

$$P(\text{“at least an } A- \text{ in IE 300"}) = P(E_1) + P(E_2) = 50\%.$$

From the three laws of probability, we also deduce that:

- $P(\overline{E}) = 1 - P(E)$.
- $P(\emptyset) = 0$.
- If $E_1 \subseteq E_2$, then $P(E_1) \leq P(E_2)$.

Recall that we say that one event E_1 is contained in another event E_2 and write that $E_1 \subseteq E_2$ if all outcomes that satisfy E_1 are included in E_2.

When $E_1 \subseteq E_2$

A pizza store advertises delivery in 30 minutes or less. Assume that the probability of an order being delivered in 30 minutes or less is 0.9. Then:

- the probability of an order being delivered in 15 minutes or less is at most 0.9.
- the probability of an order being delivered in 1 hour or less is at least 0.9.

An email is categorized as one of the following 2 (mutually exclusive) categories: spam or not-spam. Emails that are not-spam are also categorized as one of 3 (mutually exclusive again) categories: urgent, normal priority, and advertisements. Answer the following questions.

- If the probability of a message being spam is 0.45, then the probability of a message being not-spam is 0.55. True or False?
- What is the probability that an urgent email is spam?
- An email is urgent with probability 0.1 and an email is of normal priority with probability 0.2. Which of the following cases is true?
  1. $P(\textit{not-spam}) < 0.3$.
  2. $P(\textit{not-spam}) = 0.3$.
  3. $P(\textit{not-spam}) \geq 0.3$.

Unions and intersections of events

For any two events E_1, E_2, define $E = E_1 \cup E_2$. We then have that $E_1 \subseteq E$ and $E_2 \subseteq E$, which leads to

$$\boxed{P(E_1) \leq P(E),\ P(E_2) \leq P(E)}$$

and

$$\boxed{P(E) \leq P(E_1) + P(E_2).}$$ [4]

We can use a similar deduction for two events E_1, E_2 and $E = E_1 \cap E_2$. We have that $E \subseteq E_1$ and $E \subseteq E_2$, and get

$$\boxed{P(E) \leq P(E_1),\ P(E) \leq P(E_2).}$$

But how can we calculate $P(E_1 \cup E_2)$ exactly in the general case [5]? Let us turn back to sets and cardinalities.

[4] Recall that we already saw when the equality holds.

[5] Recall again that if E_1 and E_2 are mutually exclusive, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ by the third Kolmogorov axiom.

Deriving that $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

We have already seen how $E_1 \cup E_2$ can be written as the union of three mutually exclusive events: $E_1 \setminus E_2$, $E_1 \cap E_2$, and $E_2 \setminus E_1$. From the third Kolmogorov axiom, we have that:

$$P(E_1 \cup E_2) = P(E_1 \setminus E_2) + P(E_1 \cap E_2) + P(E_2 \setminus E_1). \quad (1)$$

Now, we note that E_1 (and E_2, respectively) can also be written as the union of two mutually exclusive events: $E_1 = (E_1 \setminus E_2) \cup (E_1 \cap E_2)$ (and $E_2 = (E_2 \setminus E_1) \cup (E_1 \cap E_2)$, respectively). This gives that

$$P(E_1) = P(E_1 \setminus E_2) + P(E_1 \cap E_2) \implies$$
$$P(E_1 \setminus E_2) = P(E_1) - P(E_1 \cap E_2). \quad (2)$$

Combining (1) and (2) gives us that:

$$\boxed{P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).}$$

Recall the grades from 3 different professors for the same class shown in a previous lecture.

Letter Grade	Professor 1	Professor 2	Professor 3	Total
A	108	20	30	158
B	44	49	46	139
C	11	15	15	41
D	0	1	8	9
Total	163	85	99	347

Assuming you call on one student out of the total 347 students, what is the probability:

1. E_1: you pick a student from Professor 1's class?
$$P(E_1) = 163/347 = 0.4697.$$
2. E_2: you pick a student who received an A in the class?
$$P(E_2) = 158/347 = 0.4553.$$
3. $E_1 \cap E_2$: you pick a student who was both in Professor 1's class and received an A in the class?
$$P(E_1 \cap E_2) = 108/347 = 0.3112.$$

How about the probability that you pick either a student from Professor 1's class or a student who received an A in the class?

Recall that there is a:

- 1% probability for a person to have been in close contact with a known COVID-19 case;
- 15% probability for a person to work as an essential worker;
- 6% probability that a person has traveled to a location with high contagion risk.

However, when estimating the probability that a person qualifies for the test, we have found that only 17% of the population does that. Based on your knowledge so far, does that make sense? How could that happen?

This last question should get us thinking about union of more than 2 events. For the next part see also the Worksheet for Lecture 3.

Deriving the probability of the union of more than 2 events

This will be filled after the actual lecture.

Deriving the probability of the union of more than 2 events (cont'd)

This will be filled after the actual lecture.

In general, for m events $E_1, E_2, \ldots, E_m$, we have:

1. Add the probabilities of the individual events.
2. Subtract the probabilities of the intersections of any two events.
3. Add the probabilities of the intersections of any three events.
4. Continue subtracting the probabilities of the intersections of any $4, 6, \ldots$ events and adding the probabilities of the intersections of any $5, 7, \ldots$, events.

Conditional probabilities

Motivation

It is common to want to recalculate our chances as more information become available or under certain conditions. For example, the probability that I miss the second leg of my flights is immediately affected by any delays I might experience the first leg of my flight. In such cases, we turn to *conditional probability*.

Definitions

Definition 2 (Conditional probability) *Conditional probability is defined as the probability that an event E_1 happens given that event E_2 has already happened: this is written as $P(E_1|E_2)$* [6].

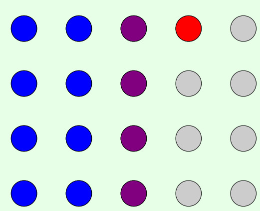
Let us think what we might need to calculate such a probability. Assume that events E_1 and E_2 (not necessarily mutually exclusive) are set to happen. We have already calculated that:

- $P(E_1) = 0.5$.
- $P(E_2) = 0.25$.
- $P(E_1 \cap E_2) = 0.2$.

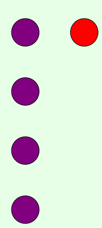
Based on the given probabilities, we can also deduce that $P(E_1 \cup E_2) = 0.5 + 0.25 - 0.2 = 0.55$, even though we do not need this result.

Changing our perception

The probability that E_1 happens is 0.5 (50%). Does this perception change if we are told that E_2 has already happened? Let us try to come up with a visual parallel to the provided probabilities. Here, we have 20 dots, out of which 12 (60%) are red and 5 (25%) are blue. 20% of them (4 in number) are both red and blue for a "purplish" color. Note that the four purple dots are both red and blue.



Had you known that E_2 has already happened, this leaves you with much fewer cases to consider!



Should our perception for the probability of E_1 change then?

[6] This is read as "the probability of E_1 given E_2" or "the probability of E_1 such that E_2 has happened".

Definition 3 (Conditional probability formula) *The conditional probability of one event E_1 conditional to event E_2 is calculated by*[7]

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}. \tag{3}$$

[7] Note that conditional probabilities are **only** for $P(E_2) > 0$.

> What is $P(E_1|E_2)$ in the previous example? How about $P(E_2|E_1)$?

Finally, note that for two mutually exclusive events E_1, E_2, the definition of conditional probabilities certainly implies that

$$P(E_2|E_1) = 0 \text{ and } P(E_1|E_2) = 0.$$

The multiplication rule for probabilities

A very straightforward rewriting of the conditional probability formula gives us a very important result. Solving for the numerator of the right hand side in (3) gives us that for any two events A, B:

$$P(A \cap B) = P(A|B) \cdot P(B). \tag{4}$$

This will come quite in handy in the next lecture.

Independence

We typically say that two entities are independent if actions of one are completely unaffected (and do not themselves affect) the actions of the other. In probability theory, we say that two events are **independent events** if knowledge that one has happened (or not) does not affect our perception for the probability of the other.

In mathematical terms, we say the following.

Definition 4 (Independent events) *Two events E_1, E_2 are independent if we have that:*

$$P(E_2|E_1) = P(E_2) \textit{ and } P(E_1|E_2) = P(E_1).$$

Equivalently, we may write that two events E_1, E_2 are independent if

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2).$$

> Think of two events from real life that are independent. Also think of two events that are clearly dependent.

> Two events are mutually exclusive. Are they independent?