



# STAT 131 - Intro to Probability Theory

## Lecture 2: Conditional Probability

Suzana Şerboi, Ph.D.  
Mathematics and Statistics, UCSC

# Thinking conditionally

Conditional probability addresses the fundamental question of how we should update our beliefs based on the evidence we observe.

You could think of every probability being conditional, whether or not it's written explicitly. For example, consider a morning when we are interested in the event  $R$  that it will rain that day.

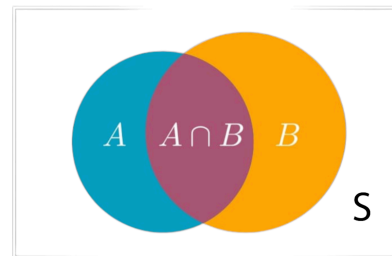
- Let  $P(R)$  represent our initial assessment of the probability of rain before looking outside.
- If we then look outside and see it is cloudy, our probability of rain should presumably increase; we denote this new probability as  $P(R|C)$  (read as “probability of  $R$  given  $C$ ”), where  $C$  is the event of being cloudy.
- When we go from  $P(R)$  to  $P(R|C)$ , we say we are “conditioning on  $C$ ”.
- Throughout the day, as we gather more information about the weather conditions, we can continuously update our probabilities.
- If we observe events  $B_1, B_2, \dots, B_n$ , then we express our updated conditional probability of rain given this evidence as  $P(R|B_1, \dots, B_n)$ .
- Ultimately, if it does start raining, our conditional probability becomes 1.

# Conditional Probability Definition

Let  $A$  and  $B$  be events with  $p(B) > 0$ . The **conditional probability** of  $A$  given  $B$ , denoted by  $p(A | B)$ , is defined as

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

Here  $A$  is the event whose uncertainty we want to update, and  $B$  is the evidence we observe (or want to treat as given).



We call  $p(A)$  the **prior probability** of  $A$  and  $p(A|B)$  the **posterior probability** of  $A$  (“prior” means before updating based on the evidence, and “posterior” means after updating based on the evidence).

See an interesting example at <https://setosa.io/conditional/>

## Example

- Standard deck: 52 cards that are shuffled.
- 2 cards are drawn randomly, one at a time without replacement.
- Let  $A$  be the event that the first card is a heart.
- Let  $B$  be the event that the second card is red.
- Find  $P(A|B)$  and  $P(B|A)$ . Are they equal?

Using the naive definition of probability and the multiplication rule:

$$P(A \cap B) = \frac{13 \cdot 25}{52 \cdot 51} = \frac{25}{204},$$

$$P(A) = \frac{1}{4} \text{ and } P(B) = \frac{1}{2}.$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{25/204}{1/2} = \frac{25}{102},$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{25/204}{1/4} = \frac{25}{51}.$$

In general **it is NOT the case that**  $p(A|B) = p(B|A)$ .

People get this confused all the time. For example, saying that there is a very high probability that an animal has four legs if it is a dog, is not the same as the probability that an animal is a dog if it has four legs.

In this case, the difference between  $p(A|B)$  and  $p(B|A)$  is obvious but there are cases where it is less obvious.

This mistake is made often enough in legal cases that it is sometimes called the prosecutor's fallacy. See Example 2.8.1 page 74.

### The Prosecutor's Fallacy



A consequence the definition of conditional probability is obtained easily by moving the denominator in the definition to the other side of the equation.

**Probability of the intersection of two events.** For any events  $A$  and  $B$  with positive probabilities,

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

Generalizing to the intersection of  $n$  events:

**Probability of the intersection of  $n$  events.** For any events  $A_1, \dots, A_n$  with  $P(A_1, A_2, \dots, A_n) > 0$ ,

$$P(A_1, \dots, A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1, A_2) \cdot \dots \cdot P(A_n|A_1, \dots, A_{n-1})$$

The commas denote intersections, e.g.,  $P(A_3|A_1, A_2) = P(A_3|A_1 \cap A_2)$ .

Note that there are  $n!$  theorems in one, since we can permute  $A_1, \dots, A_n$  in  $n!$  different ways.

Bayes' rule is an extremely famous, extremely useful result that relates  $P(A|B)$  to  $P(B|A)$ .

Bayes' rule has important implications and applications in probability and statistics, since it is so often necessary to find conditional probabilities, and often  $P(B|A)$  is much easier to find directly than  $P(A|B)$  (or vice versa).

**Bayes' rule.**

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

Let  $A$  and  $B$  be events. We may express  $B$  as  $B = (B \cap A) \cup (B \cap A^c)$ . Since  $B \cap A$  and  $B \cap A^c$  are clearly mutually exclusive, we have, by 2nd axiom of probability,

$$\begin{aligned} P(B) &= P((B \cap A) \cup (B \cap A^c)) = P(B \cap A) + P(B \cap A^c) \\ &= P(B|A)P(A) + P(B|A)P(A^c). \end{aligned}$$

Thus,

$$P(B) = P(B|A)P(A) + P(B|A) [1 - P(A)]$$

**Example.** An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. The company's statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability .4, whereas this probability decreases to .2 for a person who is not accident prone. If we assume that 30 percent of the population is accident prone, what is the probability that a new policyholder will have an accident within a year of purchasing a policy?

**Solution.** We shall obtain the desired probability by first conditioning upon whether or not the policyholder is accident prone. Let  $A_1$  denote the event that the policyholder will have an accident within a year of purchasing the policy, and let  $A$  denote the event that the policyholder is accident prone. Hence, the desired probability is given by

$$P(A_1) = P(A_1|A)P(A) + P(A_1|A^c)P(A^c) = (.4)(.3) + (.2)(.7) = .26$$

Suppose that a new policyholder has an accident within a year of purchasing a policy. What is the probability that he or she is accident prone?

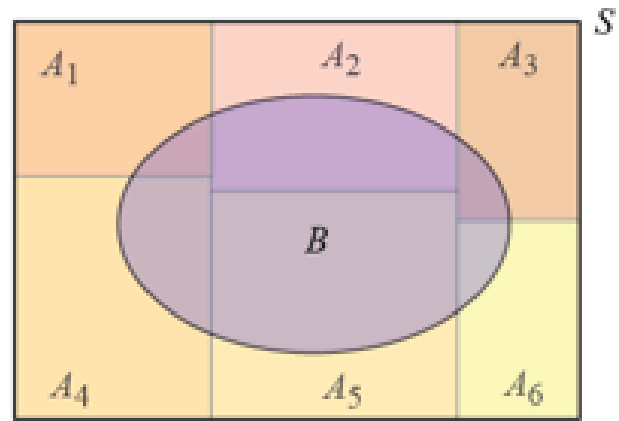
Using Bayes' Rule,

$$P(A|A_1) = \frac{P(A_1|A)P(A)}{P(A_1)} = (.3)(.4)/(0.26) = 6/13$$



**Law of total probability (LOTP).** Let  $A_1, \dots, A_n$  be a partition of the sample space  $S$  (i.e., the  $A_i$  are disjoint events and their union is  $S$ ), with  $P(A_i) > 0$  for all  $i$ . Then:

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



**Proof.** Since  $A_i$  form a partition of  $S$ , we can decompose  $B$  as  $B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$ .

As we know these pieces are disjoint, we can use the 2nd axiom of probability:

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_n)$$

Then we rewrite each intersection:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_n)P(A_n)$$

**Example (Testing for a rare disease).** A patient is tested for a rare disease (that we know affects 1% of the population). The test result is positive. Let  $D$  be the event that the patient has the disease, and  $T$  be the event that she tests positive.

Suppose the test is 95% accurate, i.e.

- $P(T|D) = 0.95$  (sensitivity or true positive rate) and
- $P(T^c|D^c) = 0.95$  (specificity or true negative rate).

What is the conditional probability that the patient has the rare disease, given that her test was positive  $P(D|T)$ ?

We can use Bayes' Rule in the first step, and then the LOTP:

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)}$$

Then we plug in the defined values:

$$P(D|T) = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \approx 0.16.$$

# Conditional probabilities

**Conditional probabilities are probabilities.** Some important results:

- Conditional probabilities are between 0 and 1.
- 1st Axiom:  $P(S|E) = 1, P(\emptyset|E) = 0$
- 2nd Axiom: If  $A_1, A_2, \dots$  are disjoint, then

$$P\left(\bigcup_{j=1}^{\infty} A_j|E\right) = \sum_{j=1}^{\infty} P(A_j|E).$$

- Complement:  $P(A^c|E) = 1 - P(A|E)$ .
- Inclusion-exclusion:  
 $P(A \cup B|E) = P(A|E) + P(B|E) - P(A \cap B|E)$ .
- Remember  $A|E$  is NOT an event. We calculate this assuming that  $E$  is already given (it has occurred).

### Proof of Axiom 1 for conditional probability:

- Fix an event  $E$ , with  $P(E) > 0$ , and for any event  $A$ , define  $\tilde{P}(A) = P(A|E)$
- $\tilde{P}(\emptyset) = P(\emptyset|E) = \frac{P(\emptyset \cap E)}{P(E)} = 0$
- $\tilde{P}(S) = P(S|E) = \frac{P(S \cap E)}{P(E)} = 1$

### Proof of Axiom 2 for conditional probability:

- Fix an event  $E$ , with  $P(E) > 0$ , and for any event  $A$ , define  $\tilde{P}(A) = P(A|E)$ .
- If we define  $A_1, A_2, \dots$  as disjoint events, then:

$$\begin{aligned}\tilde{P}(A_1 \cup A_2 \cup \dots) &= P(A_1 \cup A_2 \cup \dots | E) = \frac{P((A_1 \cap E) \cup (A_2 \cap E) \cup \dots)}{P(E)} \\ &= \frac{\sum_{j=1}^{\infty} P(A_j \cap E)}{P(E)} = \sum_{j=1}^{\infty} \tilde{P}(A_j) = \sum_{j=1}^{\infty} P(A_j | E)\end{aligned}$$

Therefore  $\tilde{P}$  satisfies the axioms of probability.

**Bayes' rule with extra conditioning.** Provided that  $P(A \cap E) > 0$  and  $P(B \cap E) > 0$ , we have

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

Note we also have

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

**LOTP with extra conditioning.** Let  $A_1, \dots, A_n$  be a partition of the sample space  $S$  (i.e., the  $A_i$  are disjoint events and their union is  $S$ ), with  $P(A_i \cap E) > 0$  for all  $i$ . Then:

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

# Independence of events

**Motivation:** Events are independent if the occurrence of one does not affect the occurrence of the other i.e.

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

**Independence of two events.** Events  $A$  and  $B$  are independent if

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

If  $P(A) > 0$  and  $P(B) > 0$ , then this is equivalent to:  $P(A|B) = P(A)$  and also equivalent to  $P(B|A) = P(B)$ .

**Example** Rolling two dice, let  $A$  be the event that the sum of the numbers is 7,  $B$  the event that the first die shows 3. These events are independent:

$$\frac{1}{36} = P(A \cap B) = P(A) \cdot P(B) = \frac{6}{36} \cdot \frac{1}{6} = \frac{1}{36}$$

If  $A$  and  $B$  are independent, then  $A$  and  $B^c$  are independent,  $A^c$  and  $B$  are independent, and  $A^c$  and  $B^c$  are independent.

**Proof.** Let  $A$  and  $B$  be independent. If  $P(A) = 0$ , then  $A$  is independent of every event, including  $B^c$ . Thus we assume  $P(A) \neq 0$ . Then

$$P(B^c|A) = 1 - P(B|A) = 1 - P(B) = P(B^c)$$

so  $A$  and  $B^c$  are independent. The other results can be proven in a similar way.

**Independence of three events.** Events  $A$ ,  $B$ , and  $C$  are said to be independent if all of the following hold:

- $P(A \cap B) = P(A)P(B)$  (pairwise independence of  $A$  and  $B$ )
- $P(A \cap C) = P(A)P(C)$  (pairwise independence of  $A$  and  $C$ )
- $P(B \cap C) = P(B)P(C)$  (pairwise independence of  $B$  and  $C$ )
- $P(A \cap B \cap C) = P(A)P(B)P(C)$ .

- You can define independence of any number of events similarly.
- Pairwise independence doesn't imply independence of more events.

# Conditional independence of events

**Conditional independence.** Events  $A$  and  $B$  are said to be conditionally independent given  $E$  if

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

Independence is a tricky concept. Some notes:

- Two events can be conditionally independent given  $E$  but not independent given  $E^c$ .
- Conditional independence doesn't imply independence (see Example 2.5.10 page 65)
- Independence doesn't imply conditional independence (see Example 2.5.11 page 66)



# Additional Practice Problems\*

1. A government agency has developed a scanner which determines whether a person is a terrorist. The scanner is fairly reliable; 99% of all scanned terrorists are identified as terrorists, and 95% of all upstanding citizens are identified as such. An informant tells the agency that exactly one passenger of 100 aboard an airplane in which you are seated is a terrorist. The agency decide to scan each passenger, and the shifty looking man sitting next to you tests positive. What are the chances that this man really is a terrorist?

2. We have a test such that  $P(T^+|HIV+) = 0.99$  (sensitivity) and  $P(T^-|HIV+) = 0.99$  (specificity). Assume the prevalence of HIV is 0.001.

a. Given the first test is positive, event  $T_1^+$ , what is the probability of being HIV positive?

b. The policy for a positive HIV test is a follow-up confirmatory test. Given the 2nd test is positive, event  $T_2^+$ , what is the probability of being HIV positive?

We can assume the new test is independent of the original test (given his disease status) and has the same sensitivity and specificity.