Principles and Practices of Data Science

Lecture 7

Melvin Ayala

Lecture 7: Bayes’ Theorem and Applications

Sections

1. Probability of Events
2. Probability of Negation of Events
3. Probability of Intersection of Events
4. Probability of Union of Events
5. Independent Events
6. Independence and Probability
7. Introduction to Conditional Probabilities
8. Bayes’ Theorem
9. Real-World Applications

1. Probability of Events

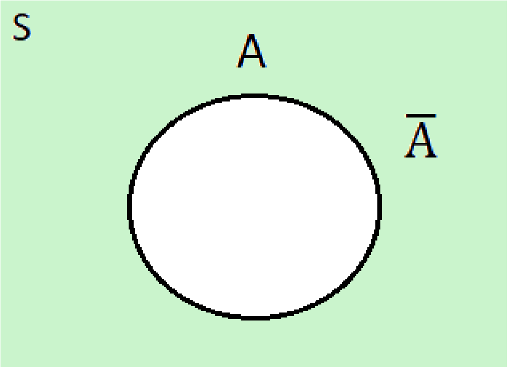
Brief Definition

- the extent to which an event is likely to occur, measured by the ratio of the favorable cases to the whole number of cases possible.

$$Probability\ of\ an\ event = \frac{Number\ of\ avorable\ cases}{Number\ of\ possible\ cases}$$

2. Probability of Negation of Events

Consider a sample space with an event A.



- the probability of any event in S to be a member of the sample space S is always 1.

  P(S) = 1

- the negation of A, denoted $\overline{\mathrm{A}}$, consists of all the cases in S that are not included in A.
- the probability of $\overline{\mathrm{A}}$ is the complement of the probability of A.

  $$P(\bar{A}) + P(A) = 1$$

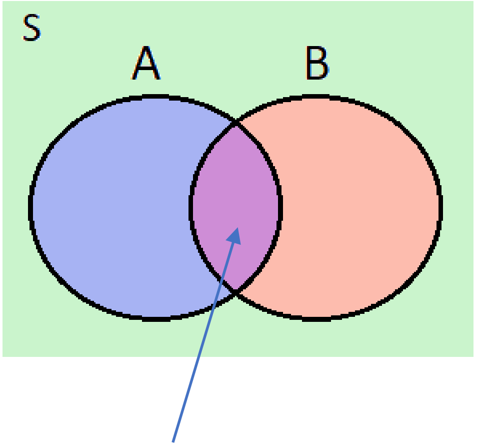
  i.e.,

  $$P(\bar{A}) = 1 - P(A)$$

3. Probability of Intersection of Events

Consider a sample space with events A and B.

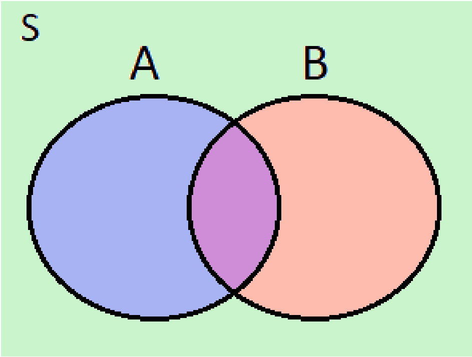
- the shaded section of the Venn diagram below is the outcomes shared by events A and B.
- it is called intersection of events A and B.
- notation: $A \cap B$.
- note that $A \cap B$ is equivalent to $B \cap A$.
- for independent events: $P(A \cap B) = P(A)\, P(B)$
- for dependent events: use the Bayes theorem



$P(A \cap B)$

4. Probability of Union of Events

Consider a sample space with events A and B.



- number of outcomes in a union of events?

  $P(A \cup B) \geq P(A) + P(B)$

- outcomes in event A + outcomes in event B → double counting

- how to count?
  - count the number of outcomes in each event separately
  - subtract the number of outcomes shared by both events

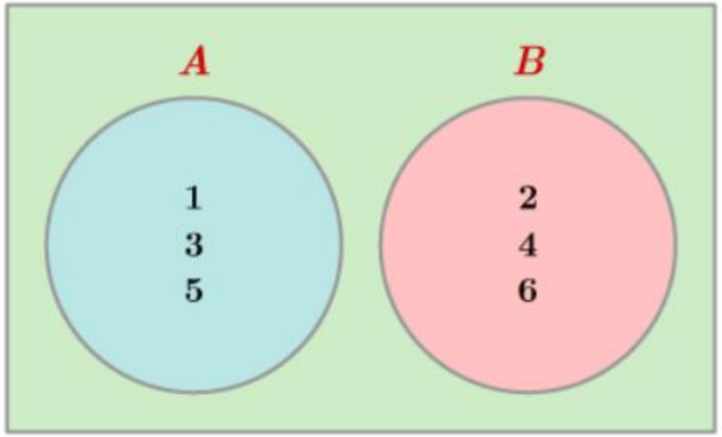
- generalizing to probability:

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

5. Independent Events

Disjoint Events

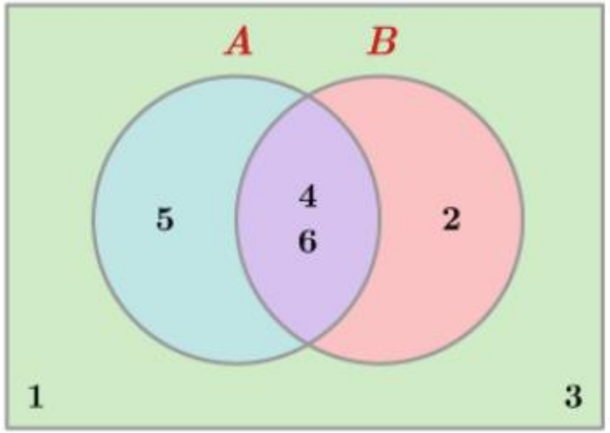
Experiment: Rolling a single die
Event A: Get an odd number
Event B: Get an even number



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

Overlapping Events

Experiment: Rolling a single die
Event A: Get a number over 3
Event B: Get an even number



$(A \cap B) \neq \varnothing$
(dependent)

Definition:

Two events are independent → occurrence of one event does not affect the chances of the occurrence of the other event.

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

Independent Events (Cont.)

When are two events or random variables (statistically) independent?

- when the occurrence of one event (or random variable) does not affect the chances of the occurrence of the other event (or random variable)
- if knowing the value of one of them does not change the probabilities for the other one.

Mathematical Formulation:

If
P(A and B) = P(A)P(B)),
then A and B are independent.

Using conditional probability notation:

P(B=b|A=a) = P(B=b)
P(A=a|B=b) = P(A=a)
for all a,b

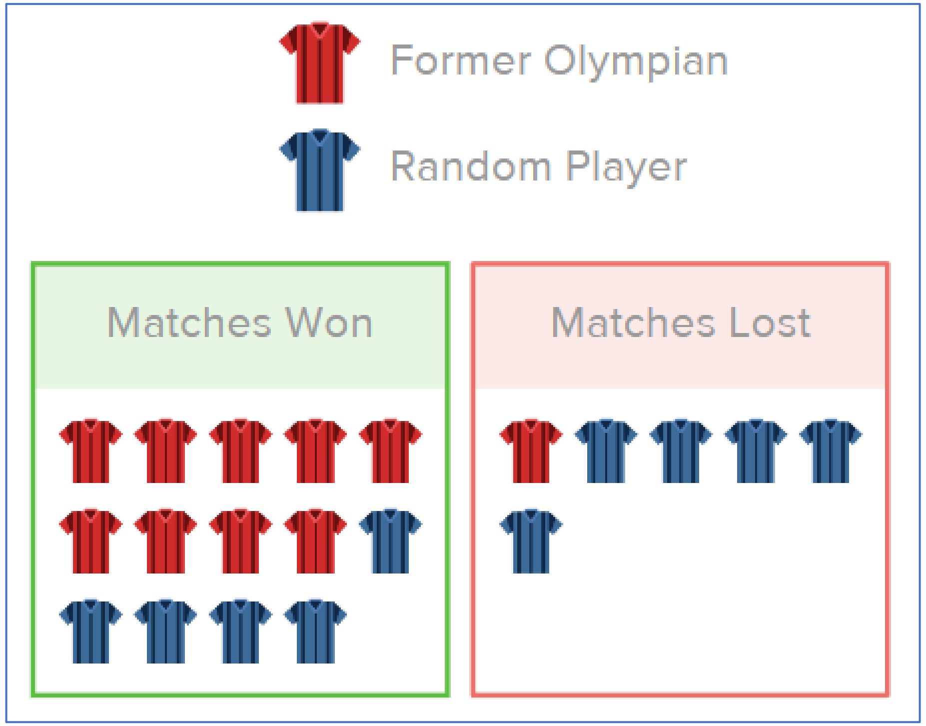
6. Independence and Probability

Independence and Probability

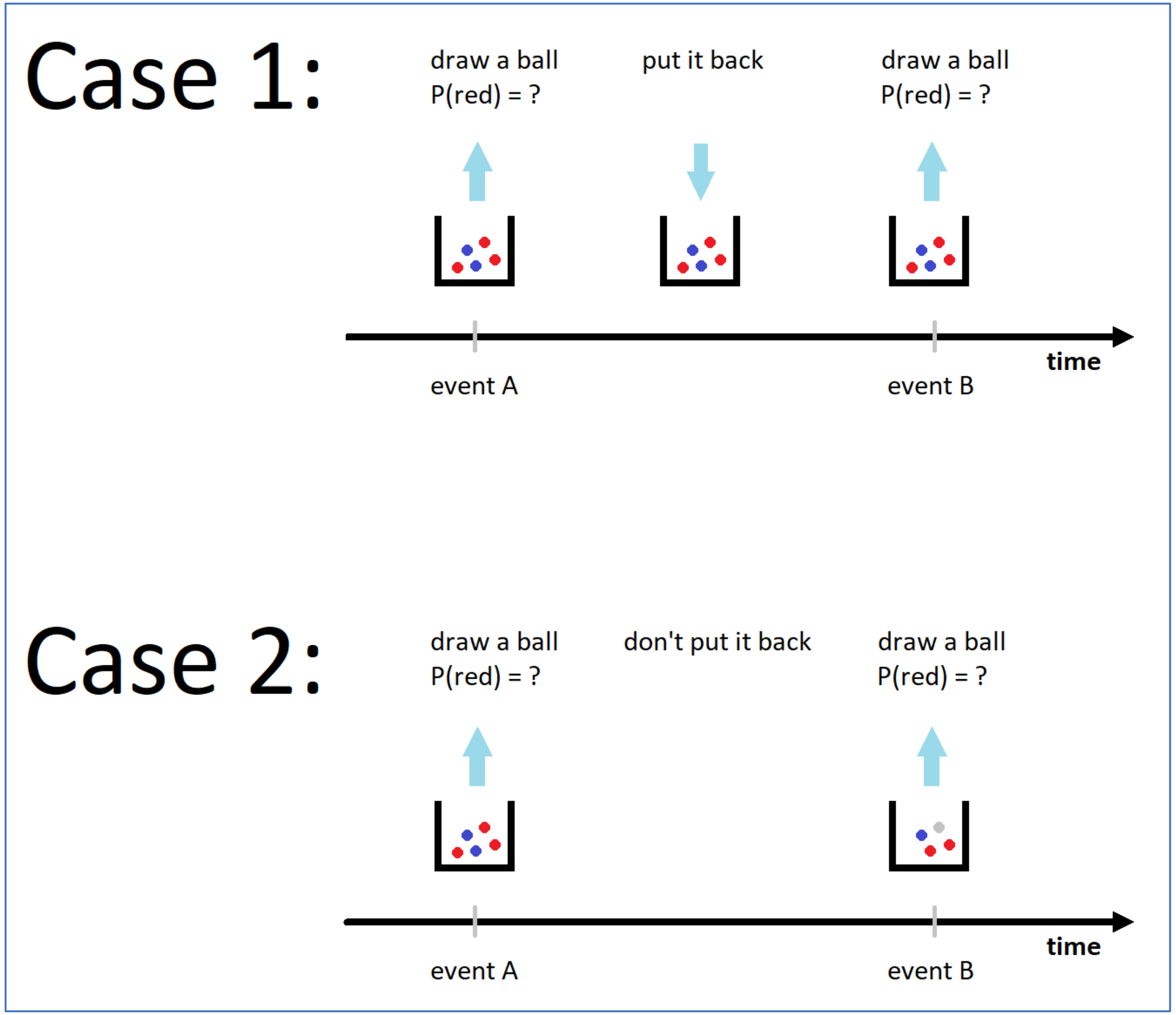
Conditional Probability:

- a probability where additional information is known.
- probabilities below are different:
  - P(team scoring better when coach randomly hired)
  - P(team scoring better when coach former olympian)
- the additional information (coach being a former olympian) changes the probability.
- if the additional information does not ultimately change the probability, then the two events are independent.

Probability of Winning a Match



Example of Dependent and Independent Events

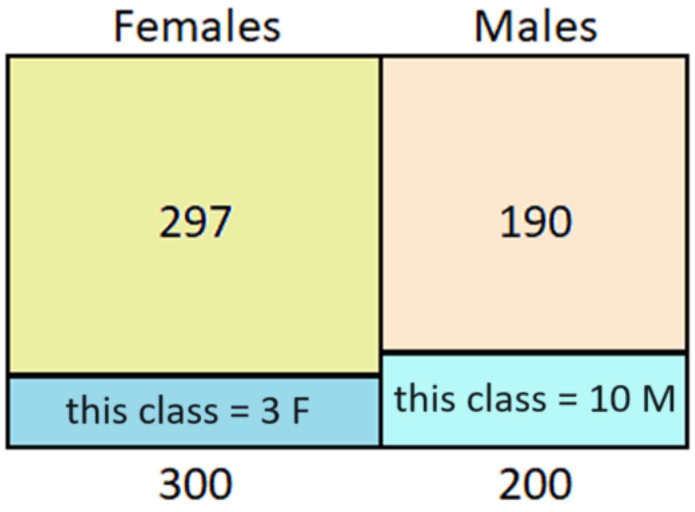


← events A and B are independent.

← events A and B are dependent.

Can you guess why?

7. Introduction to Conditional Probabilities



Suppose we have 500 students in the building taking classes now.
Suppose of all these students, 300 are females and 200 are males.
Suppose we have 3 females and 10 males in class now.

Let

P(F)	probability of a student to be female.
P(M)	probability of a student to be male.
P(this class)	probability of a student to be in our class now.
P(not this class)	probability of a student not to be in our class now.
P(F | this class)	probability of a student to be female given that the student is in our class now.
P(M | this class)	probability of a student to be male given that the student is in our class now.
P(this class | F):	probability of a student to be in our class now given that the student is female.
P(this class | M)	probability of a student to be in our class now given that the student is male.

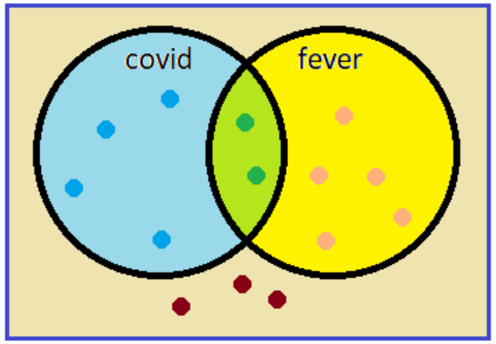
then:
P(F) = (297 + 3) / 500 = 0.6
P(M) = (190 + 10) / 500 = 0.4
P(this class) = (3 + 10) / 500 = 0.026
P(not this class) = (297 + 190) / 500 = 0.974

We can find:
P(F | this class) = ?
P(M |this class) = ?
P(this class | F) = ?
P(this class | M) = ?

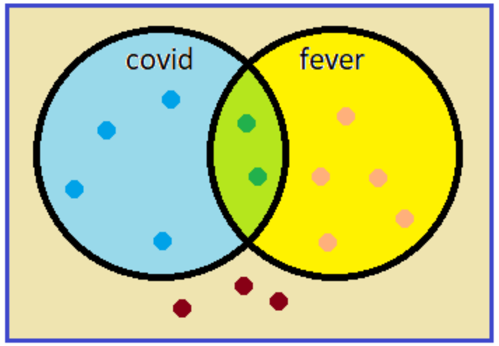
A Conditional Probability Example (Fever-Covid)

Suppose we have 14 patients in a clinic at a given day and time, such that:

- 6 patients are diagnosed with covid.
- 7 patients have fever.
- only 2 patients with covid have fever.
- 3 patients have neither covid nor fever.



		covid		
		Yes	No	row total
fever	Yes	2	5	7
	No	4	3	7
col total		6	8	



		covid		
		Yes	No	row total
fever	Yes	2 p=2/14	5 p=5/14	2+5=7 p=7/14
	No	4 p =4/14	3 p=3/14	4+3=7 p=7/14
col total		2+4=6 p=6/14	5+3=8 p=8/14	

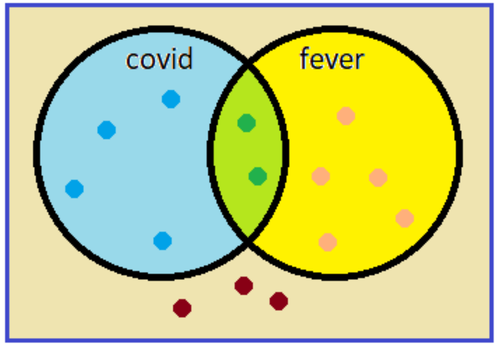
Probabilities:
P(fever & covid) = 2/14
P(fever & no covid) = 5/14
P(no fever & covid) = 4/14
P(no fever & no covid) = 3/14

More probabilities:
P(fever) = (2+5)/14
P(no fever) = (4+3)/14
P(covid) = (2+4)/14
P(no covid) = (5+3)/14

A Conditional Probability Example (Fever-Covid)(Cont.)

Let's calculate the conditional probability that a patient might not have covid but has fever,
given that we already know that the patient has fever.

P(no covid & fever | fever) = ?



		covid		
		Yes	No	row total
fever	Yes	2 p=2/14	5 p=5/14	2+5=7 p=7/14
	No	4 p =4/14	3 p=3/14	4+3=7 p=7/14
col total		2+4=6 p=6/14	5+3=8 p=8/14	

Solution:

- divide the 5 patients who have fever but do not have covid by the 7 people who have fever (i.e., scaling by fever)

P(no covid & fever | fever) = 5/(2+5) = 5/7 = 0.7143

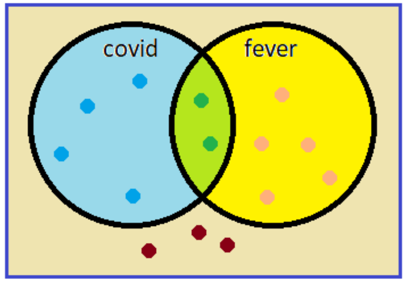
Just for fun: divide both the numerator and the denominator by the total number of patients:

$$\mathrm{P(no\ covid\ \&\ fever\ |\ fever)} = \frac{5}{2+5} = \frac{\frac{5}{14}}{\frac{2+5}{14}} = \frac{\mathrm{P(no\ covid\ \&\ fever)}}{\mathrm{P(fever)}}$$

Eliminate redundancy:

$$\mathrm{P(no\ covid\ |\ fever)} = \frac{\mathrm{P(no\ covid\ \&\ fever)}}{\mathrm{P(fever)}}$$

A Conditional Probability Example (Fever-Covid)(Cont.)



		covid		
		Yes	No	row total
fever	Yes	2 p=2/14	5 p=5/14	2+5=7 p=7/14
	No	4 p =4/14	3 p=3/14	4+3=7 p=7/14
col total		2+4=6 p=6/14	5+3=8 p=8/14	

P(no covid & fever | fever) = ? → scale by fever

$$P(\text{no covid \& fever} \mid \text{fever}) = \frac{N(\text{no covid \& fever})}{N(\text{fever})} = \frac{5}{7}$$

Let's divide both numerator and denominator by the total number of patients (14):

$$P(\text{no covid \& fever} \mid \text{fever}) = \frac{\frac{N(\text{no covid \& fever})}{14}}{\frac{N(\text{fever})}{14}} = \frac{\frac{5}{14}}{\frac{7}{14}} = \frac{P(\text{no covid \& fever})}{P(\text{fever})} = \frac{5}{7}$$

Eliminating redundancy:

$$P(\text{no covid \& fever} \mid \text{fever}) = P(\text{no covid} \mid \text{fever})$$

$$P(\text{no covid} \mid \text{fever}) = \frac{P(\text{no covid \& fever})}{P(\text{fever})} = \frac{5}{7}$$

P(no covid & fever | no covid) = ? → scale by no covid

$$P(\text{no covid \& fever} \mid \text{no covid}) = \frac{N(\text{no covid \& fever})}{N(\text{no covid})} = \frac{5}{8}$$

Let's divide both numerator and denominator by the total number of patients (14):

$$P(\text{no covid \& fever} \mid \text{no covid}) = \frac{\frac{N(\text{no covid \& fever})}{14}}{\frac{N(\text{no covid})}{14}} = \frac{\frac{5}{14}}{\frac{8}{14}} = \frac{P(\text{no covid \& fever})}{P(\text{no covid})} = \frac{5}{8}$$

Eliminating redundancy:

$$P(\text{no covid \& fever} \mid \text{no covid}) = P(\text{fever} \mid \text{no covid})$$

$$P(\text{fever} \mid \text{no covid}) = \frac{P(\text{no covid \& fever})}{P(\text{no covid})} = \frac{5}{8}$$

$$P(\text{no covid} \mid \text{fever}) \cdot P(\text{fever}) = P(\text{fever} \mid \text{no covid}) \cdot P(\text{no covid})$$

$$\frac{5}{7} \cdot \frac{7}{14} = \frac{5}{8} \cdot \frac{8}{14}$$

8. Bayes’ Theorem

What Is Bayes' Theorem?

- Bayes' Theorem, named after 18th-century British mathematician Thomas Bayes, is a mathematical formula for determining conditional probability.
- describes the probability of an event, based on prior knowledge of conditions that might be related to the event
- conditional probability is the likelihood of an outcome occurring, based on a previous outcome having occurred in similar circumstances.
- Bayes' theorem provides a way to revise existing predictions or theories (update probabilities) given new or additional evidence.

Thomas Bayes (1701– 1761)

Bayes' Theorem:

- a mathematical formula for determining conditional probability

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

where:

P(A) = probability of A occurring
P(B) = probability of B occurring
P(A|B) = probability of A given B
P(B|A) = probability of B given A
P(A∩B) = probability of both A and B occurring

Example 1: Dangerous Fire and Smoke

We have the following statistics:

- dangerous fires are rare (1%)
- smoke is fairly common (10%) due to barbecues
- 90% of dangerous fires make smoke

What is the probability of a dangerous fire when there is smoke?

Solution:

We have:

P(S|DF) = 0.9	probability to find smoke given that we see a dangerous fire
P(S) = 0.1	probability to find smoke
P(DF) = 0.01	probability to find a dangerous fire
P(DF|S) = ?	probability to find a dangerous fire given that we see smoke

Using the Bayes' theorem, we have: $P(DF|S)\,P(S) = P(S|DF)\,P(DF)$

then:

$$P(DF|S) = P(S|DF)\,P(DF)\,/\,P(S)$$
$$= 0.9\,(0.01)\,/\,0.1$$
$$= 0.09$$

Example 2: Picnic, Clouds and Rain

You are planning a picnic today, but the morning is cloudy.
We also know that:

- 50% of all rainy days start off cloudy!
- cloudy mornings are common (about 40% of days start cloudy)
- this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)

What is the chance of rain during the day?

Solution:

We have:

P(C|R) = 0.5	probability to have clouds given that it rains during the day
P(C) = 0.4	probability to have a cloudy day
P(R) = 0.1	probability to have a rainy day
P(R|C) = ?	probability to have a rainy day given that we see clouds

Using the Bayes' theorem, we have: $P(R|C)\,P(C) = P(C|R)\,P(R$
then:

$$P(R|C) = P(C|R)\,P(R) / P(C)$$
$$= 0.5\,(0.1) / 0.4$$
$$= 0.125$$

Example 3: The Cookie Problem

- **suppose there are two bowls of cookies:**
  - bowl 1: 30 vanilla cookies and 10 chocolate cookies.
  - bowl 2: 20 of each.

- **you choose one of the bowls at random and select a cookie at random:**
  - suppose the cookie is vanilla.

- **question:**
  - what is the probability that it came from bowl 1?

- this is a conditional probability.
- we want P(bowl 1 | vanilla), but it is not obvious how to compute it.
- a different question would be the probability of a vanilla cookie given Bowl 1
  - that would be easy: P(vanilla | bowl 1) = 30/40

- P(A|B) is not the same as P(B|A)
- there is a way to get from one to the other: Bayes' theorem.

Solution to the cookie problem:

data:
bowl 1: 30 vanilla cookies and 10 chocolate cookies.
bowl 2: 20 of each.
vanilla cookie chosen.

question:
what is the probability that it came from bowl 1? ← hypothesis

Bayes' theorem:

$$P(bowl\ 1\,|\,vanilla) = \frac{P(bowl\ 1)\ P(vanilla\,|\,bowl\ 1)}{P(vanilla)}$$

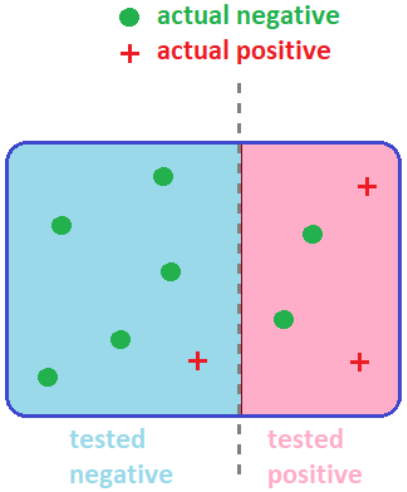
P(bowl 1) = probability that we chose bowl 1, unconditioned by what kind of cookie we got. P(bowl 1) = 1/2.
P(vanilla | bowl 1): probability of getting a vanilla cookie from bowl 1, which is 30/40 = 3/4.
P(vanilla): This is the probability of drawing a vanilla cookie from either bowl. P(vanilla) = 50/80 = 5/8.

$$P(bowl\ 1\,|\,vanilla) = \frac{P(bowl\ 1)\ P(vanilla\,|\,bowl\ 1)}{P(vanilla)} = \frac{(1/2) * (30/40)}{50/80} = \frac{3}{5}$$

vanilla cookie is evidence in favor of the hypothesis that we chose bowl 1, because vanilla cookies are more likely to come from bowl 1.

Example 4: Medical Tests

Given the following confusion matrix obtained from a medical test:



		Test Result		
		Neg	Pos	totals
Actual	Neg	5	2	7
	Pos	1	2	3
	totals	6	4	

		Test Result		
		Neg	Pos	totals
Actual	Neg	TN	FP	7
	Pos	FN	TP	3
	totals	6	4	

$N_{actualneg} = 7$
$N_{actualpos} = 3$
$N_{testneg} = 6$
$N_{testpos} = 4$

P(TestNeg) = 6/10
P(TestPos) = 4/10

1. Compute:

P(ActualNeg | TestNeg)
P(ActualNeg | TestPos)

Solution:

P(ActualNeg | TestNeg) = N(ActualNeg & TestNeg) / N(TestNeg)
= P(ActualNeg & TestNeg) / P(TestNeg)
= 5 / 6

P(ActualNeg | TestPos) = N(ActualNeg & TestPos) / N(TestPos)
= P(ActualNeg & TestPos) / P(TestPos)
= 2 / 4

2. Compute P(ActualNeg):

Solution:

P(ActualNeg) = P(ActualNeg | TestNeg) P(TestNeg) +
P(ActualNeg | TestPos) P(TestPos)
= (5/6) (6/10) + (2/4) (4/10)
= 7/10

But a faster way would have been:
P(ActualNeg) = $N_{actualneg}$ / $N_{totalpatients}$ = 7/10

Exercises:

1. What is the equation for P(ActualPos | TestNeg) ?
2. What is the equation for P(ActualPos | TestPos) ?
3. What is the equation for P(ActualPos) ?

Bayes’ Theorem and Artificial Neural Networks

Bayesian Networks:

- Bayes’ theorem can also be applied to Artificial Neural Networks.
- called Bayes networks, Bayes nets, belief networks, or decision networks.
- each node corresponds to a random variable .
- each edge represents the conditional probability for the corresponding random variables.
- Bayesian networks are ideal for taking an event that occurred and predicting the likelihood that any one of several possible known causes was the contributing factor.

Example:

- a Bayesian network could represent the probabilistic relationships between diseases and symptoms.
- given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

Naive Bayesian Networks:

- a restricted/constrained form of a general Bayesian network.
- enforce the constraint that the class node should have no parents and that the nodes corresponding to the attribute variables should have no edges between them.

9. Real-World Applications

Test Accuracy

Example:

A certain test for allergy is said to be 90% accurate.
What does this have to do with conditional probability?

Consider four groups of people:

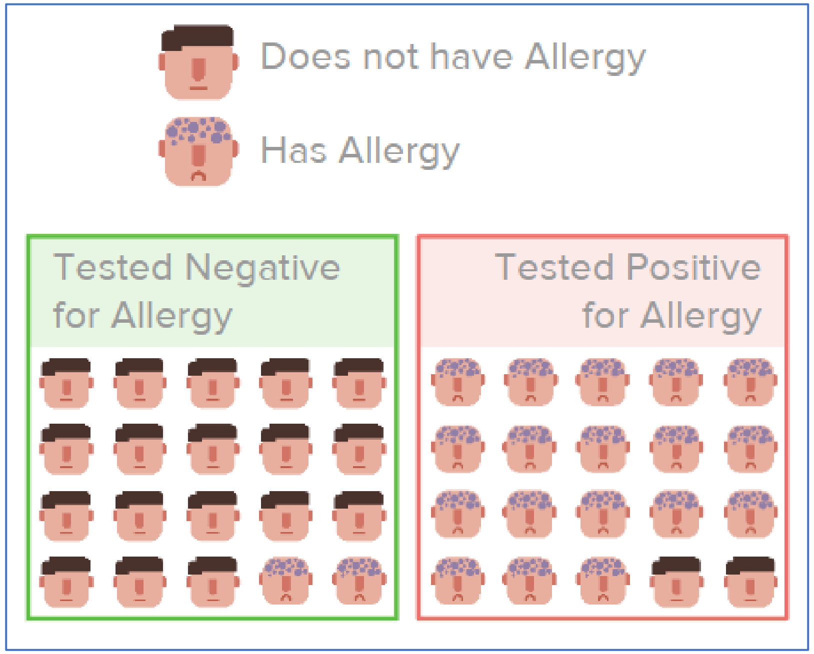
- people with the allergy who test positive for the allergy (true positive).
- people with the allergy who test negative for the allergy (false negative).
- people without the allergy who test positive for the allergy (false positive).
- people without the allergy who test negative for the allergy (true negative).

If a test is 90% accurate, it implies that:

- if a person has the allergy, 90% of the time they will receive a positive test result.
  P(positive|allergy) = 90%

- if a person does not have the allergy, 90% of the time they will receive a negative test result.
  P(negative|no allergy) = 90%

Accuracy Test for Allergy

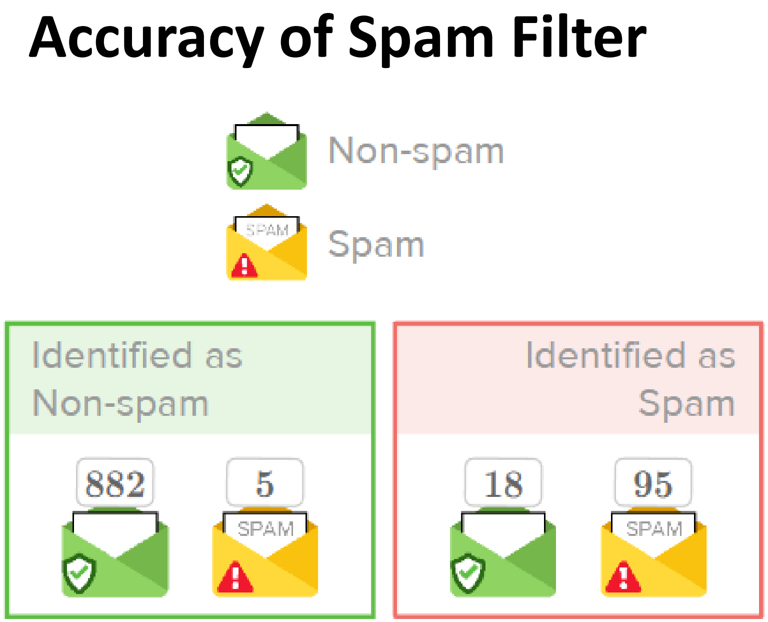


Real-World Applications (Cont.)

Spam Emails

Example:

- 10% of the emails that a person receives are spam emails.
- spam filter catches spam 95% of the time.
- spam filter misidentifies non-spam as spam 2% of the time.



Let:
A: event that an email is spam.
B: event that the spam filter identifies the email as spam.

P(A) is the probability that a random email is spam.
P(A|B) is the probability that a spam email gets identified as spam.
P(B'|A') is the probability that a non-spam email does not get identified as spam.

Find P(A): P(A) = 10%

Find P(B|A): P(B|A) = 95%

Find P(B'|A'): P(B'|A') = 98%

End of Lecture