Unit 2: Probability and distributions

1. Probability and conditional probability

GOVT 3990 - Spring 2018

Cornell University

Outline

1. Main ideas
   1. Disjoint and independent do not mean the same thing
   2. Application of the addition rule depends on disjointness of events
   3. Bayes' theorem works for all types of events
   4. Probability trees are useful for conditional probability calculations
   5. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate
   6. Posterior probability and p-value do not mean the same thing
2. Summary

Outline

1. Main ideas

    1. Disjoint and independent do not mean the same thing
    2. Application of the addition rule depends on disjointness of events
    3. Bayes' theorem works for all types of events
    4. Probability trees are useful for conditional probability calculations
    5. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate
    6. Posterior probability and p-value do not mean the same thing

2. Summary

1. Disjoint and independent do not mean the same thing

- *Disjoint (mutually exclusive) events* cannot happen at the same time
  - A voter cannot register as a Democrat and a Republican at the same time
  - But they might be a Republican and a Moderate at the same time – *non-disjoint events*
  - For disjoint A and B: $P(A\ and\ B) = 0$

1. Disjoint and independent do not mean the same thing

- *Disjoint (mutually exclusive) events* cannot happen at the same time
  - A voter cannot register as a Democrat and a Republican at the same time
  - But they might be a Republican and a Moderate at the same time – *non-disjoint events*
  - For disjoint A and B: $P(A \ and \ B) = 0$
- If A and B are *independent events*, having information on A does not tell us anything about B (and vice versa)
  - If A and B are independent:
    - $P(A \mid B) = P(A)$
    - $P(A \ and \ B) = P(A) \times P(B)$

Outline

1. Main ideas
    1. Disjoint and independent do not mean the same thing
    2. Application of the addition rule depends on disjointness of events
    3. Bayes' theorem works for all types of events
    4. Probability trees are useful for conditional probability calculations
    5. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate
    6. Posterior probability and p-value do not mean the same thing
2. Summary

2. Application of the addition rule depends on disjointness of events

- *General addition rule:* P(A or B) = P(A) + P(B) - P(A and B)
- A or B = either A or B or both

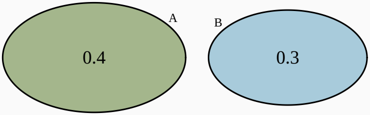
- *General addition rule:* P(A or B) = P(A) + P(B) - P(A and B)
- A or B = either A or B or both

disjoint events:

P(A or B)

= P(A) + P(B) - P(A and B)

= 0.4 + 0.3 - 0 = 0.7



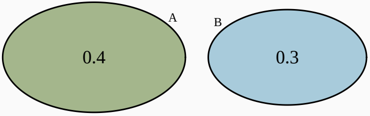
- *General addition rule:* P(A or B) = P(A) + P(B) - P(A and B)
- A or B = either A or B or both

disjoint events:

P(A or B)

= P(A) + P(B) - P(A and B)

= 0.4 + 0.3 - 0 = 0.7

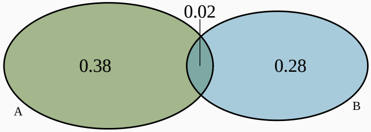


non-disjoint events:

P(A or B)

= P(A) + P(B) - P(A and B)

= 0.4 + 0.3 - 0.02 = 0.68



Outline

1. Main ideas

    1. Disjoint and independent do not mean the same thing

    2. Application of the addition rule depends on disjointness of events

    3. Bayes' theorem works for all types of events

    4. Probability trees are useful for conditional probability calculations

    5. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

    6. Posterior probability and p-value do not mean the same thing

2. Summary

3. Bayes' theorem works for all types of events

- *Bayes' theorem:* $P(A \mid B) = \frac{P(A \;\; and \;\; B)}{P(B)}$

3. Bayes' theorem works for all types of events

- *Bayes' theorem:* $P(A \mid B) = \frac{P(A\ \ and\ \ B)}{P(B)}$
- … can be rewritten as: $P(A\ and\ B) = P(A \mid B) \times P(B)$

3. Bayes' theorem works for all types of events

- *Bayes' theorem:* $P(A \mid B) = \frac{P(A \ and \ B)}{P(B)}$
- ... can be rewritten as: $P(A \ and \ B) = P(A \mid B) \times P(B)$

disjoint events:

- We know P(A | B) = 0, since if B happened A could not have happened

- *Bayes' theorem:* $P(A \mid B) = \frac{P(A \ and \ B)}{P(B)}$
- ... can be rewritten as: $P(A \ and \ B) = P(A \mid B) \times P(B)$

disjoint events:

- We know P(A | B) = 0, since if B happened A could not have happened
- P(A and B) = P(A | B) $\times$ P(B)

- *Bayes' theorem:* $P(A \mid B) = \frac{P(A \ and \ B)}{P(B)}$
- ... can be rewritten as: $P(A \ and \ B) = P(A \mid B) \times P(B)$

disjoint events:

- We know P(A | B) = 0, since if B happened A could not have happened
- P(A and B)
  = P(A | B) $\times$ P(B)
  = *0* $\times$ P(B) = 0

3. Bayes' theorem works for all types of events

- *Bayes' theorem:* $P(A \mid B) = \frac{P(A \ and \ B)}{P(B)}$
- ... can be rewritten as: $P(A \ and \ B) = P(A \mid B) \times P(B)$

disjoint events:

- We know P(A | B) = 0, since if B happened A could not have happened
- P(A and B)
  = P(A | B) $\times$ P(B)
  = *0* $\times$ P(B) = 0

independent events:

- We know P(A | B) = P(A), since knowing B doesn't tell us anything about A

- *Bayes' theorem:* $P(A \mid B) = \frac{P(A \ and \ B)}{P(B)}$
- ... can be rewritten as: $P(A \ and \ B) = P(A \mid B) \times P(B)$

disjoint events:

- We know P(A | B) = 0, since if B happened A could not have happened
- P(A and B)
  = P(A | B) $\times$ P(B)
  = *0* $\times$ P(B) = 0

independent events:

- We know P(A | B) = P(A), since knowing B doesn't tell us anything about A
- P(A and B)
  = P(A | B) $\times$ P(B)

- *Bayes' theorem:* $P(A \mid B) = \frac{P(A \ and \ B)}{P(B)}$
- ... can be rewritten as: $P(A \ and \ B) = P(A \mid B) \times P(B)$

disjoint events:

- We know P(A | B) = 0, since if B happened A could not have happened
- P(A and B)
  = P(A | B) $\times$ P(B)
  = *0* $\times$ P(B) = 0

independent events:

- We know P(A | B) = P(A), since knowing B doesn't tell us anything about A
- P(A and B)
  = P(A | B) $\times$ P(B)
  = *P(A)* $\times$ P(B)

Application exercise: 2.1 Probability and conditional probability

Outline

1. Main ideas
    1. Disjoint and independent do not mean the same thing
    2. Application of the addition rule depends on disjointness of events
    3. Bayes' theorem works for all types of events
    4. Probability trees are useful for conditional probability calculations
    5. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate
    6. Posterior probability and p-value do not mean the same thing
2. Summary

1. Probability trees are useful for conditional probability calculations

- Probability trees are useful for organizing information in conditional probability calculations
- They're especially useful in cases where you know P(A | B), along with some other information, and you're asked for P(B | A)

Outline

1. Main ideas
    1. Disjoint and independent do not mean the same thing
    2. Application of the addition rule depends on disjointness of events
    3. Bayes' theorem works for all types of events
    4. Probability trees are useful for conditional probability calculations
    5. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate
    6. Posterior probability and p-value do not mean the same thing
2. Summary

2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

- In Bayesian inference, probabilities are at times interpreted as **degrees of belief**.

2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

- In Bayesian inference, probabilities are at times interpreted as **degrees of belief**.
- You start with a set of **prior beliefs** (or prior probabilities).

2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

- In Bayesian inference, probabilities are at times interpreted as **degrees of belief**.
- You start with a set of **prior beliefs** (or prior probabilities).
- You observe some data.

2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

- In Bayesian inference, probabilities are at times interpreted as **degrees of belief**.
- You start with a set of **prior beliefs** (or prior probabilities).
- You observe some data.
- Based on that data, you update your beliefs.

2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

- In Bayesian inference, probabilities are at times interpreted as **degrees of belief**.
- You start with a set of **prior beliefs** (or prior probabilities).
- You observe some data.
- Based on that data, you update your beliefs.
- These new beliefs are called **posterior beliefs** (or posterior probabilities), because they are **post**-data.

2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

- In Bayesian inference, probabilities are at times interpreted as **degrees of belief**.
- You start with a set of **prior beliefs** (or prior probabilities).
- You observe some data.
- Based on that data, you update your beliefs.
- These new beliefs are called **posterior beliefs** (or posterior probabilities), because they are **post**-data.
- You can iterate this process.

Dice game

We'll play a game to demonstrate this approach:

- Two dice: 6-sided and 12-sided
  - I keep one die on the left and one die on the right

We'll play a game to demonstrate this approach:

- Two dice: 6-sided and 12-sided
  - I keep one die on the left and one die on the right
- The "good die" is the 12-sided die.

We'll play a game to demonstrate this approach:

- Two dice: 6-sided and 12-sided
  - I keep one die on the left and one die on the right
- The "good die" is the 12-sided die.
- Ultimate goal: come to a class consensus about whether the die on the left or the die on the right is the "good die"

We'll play a game to demonstrate this approach:

- Two dice: 6-sided and 12-sided
  - I keep one die on the left and one die on the right
- The "good die" is the 12-sided die.
- Ultimate goal: come to a class consensus about whether the die on the left or the die on the right is the "good die"
- We will start with priors, collect data, and calculate posteriors, and make a decision or iterate until we're ready to make a decision

Prior probabilities

- At each roll I tell you whether you won or not (win = ≥ 4)
  - P(win | 6-sided die) = 0.5 $\rightarrow$ bad die
  - P(win | 12-sided die) = 0.75 $\rightarrow$ good die

Prior probabilities

- At each roll I tell you whether you won or not (win = ≥ 4)
  - P(win | 6-sided die) = 0.5 $\rightarrow$ bad die
  - P(win | 12-sided die) = 0.75 $\rightarrow$ good die
- The two competing claims are

  H_1: Good die is on left

  H_2: Good die is on right

Prior probabilities

- At each roll I tell you whether you won or not (win = ≥ 4)
  - P(win | 6-sided die) = 0.5 $\rightarrow$ bad die
  - P(win | 12-sided die) = 0.75 $\rightarrow$ good die
- The two competing claims are

  H_1: Good die is on left

  H_2: Good die is on right
- Since initially you have no idea which is true, you can assign equal *prior probabilities* to the hypotheses

  P(H_1 is true) = 0.5

  P(H_2 is true) = 0.5

Rules of the game

- You won't know which die I'm holding in which hand, left (L) or right (R). left = YOUR left
- You pick die (L or R), I roll it, and I tell you if you win or not, where winning is getting a number ≥ 4. If you win, you get a piece of candy. If you lose, I get to keep the candy.
- We'll play this multiple times with different contestants.
- I will not swap the sides the dice are on at any point.
- You get to pick how long you want play, but there are costs associated with playing longer.

	Truth	
Decision	L good, R bad	L bad, R good
Pick L	*You get candy!*	*You lose all the candy :(*
Pick R	*You lose all the candy :(*	*You get candy!*

Sampling isn't free!

At each trial you risk losing pieces of candy if you lose (the die comes up < 4). Too many trials means you won't have much candy left. And if we spend too much class time and we may not get through all the material.

	Choice (L or R)	Result (win or loss)
Roll 1		
Roll 2		
Roll 3		
Roll 4		
Roll 5		
Roll 6		
Roll 7		
...		

What is your decision? How did you make this decision?

Posterior probability

- *Posterior probability* is the probability of the hypothesis given the observed data: P(hypothesis | data)

Posterior probability

- *Posterior probability* is the probability of the hypothesis given the observed data: P(hypothesis | data)
- Using Bayes' theorem

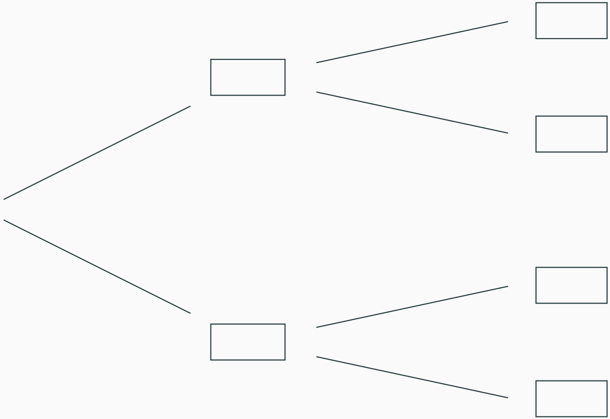
$$P(hypothesis \mid data) \quad = \quad \frac{P(hypothesis\ and\ data)}{P(data)}$$

Posterior probability

- *Posterior probability* is the probability of the hypothesis given the observed data: P(hypothesis | data)
- Using Bayes' theorem

$$
\begin{aligned}
P(hypothesis \mid data) &= \frac{P(hypothesis\ and\ data)}{P(data)} \\
&= \frac{P(data \mid hypothesis) \times P(hypothesis)}{P(data)}
\end{aligned}
$$

Calculate the posterior probability for the hypothesis chosen in the first roll, and discuss how this might influence your decision for the next roll.



Most companies drug test their employees before they start employment, and sometimes regularly during their employment as well. Suppose that a drug test for an illegal drugs is 97% accurate in the case of a user of that drug, and 92% accurate in the case of a non-user for that drug. Suppose also that 5% of the entire population uses that drug.

Most companies drug test their employees before they start employment, and sometimes regularly during their employment as well. Suppose that a drug test for an illegal drugs is 97% accurate in the case of a user of that drug, and 92% accurate in the case of a non-user for that drug. Suppose also that 5% of the entire population uses that drug.

- You are the hiring manager at a company that drug tests their employees. You have recently decided to hire a new employee. What is the prior probability that this employee is a user of this drug? (You may assume that this prospective employee is a randomly drawn person from the population.)

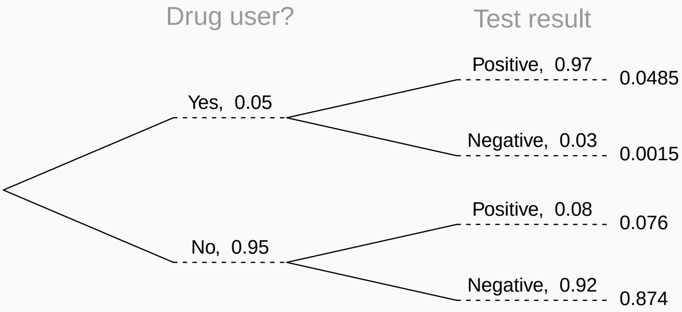
Most companies drug test their employees before they start employment, and sometimes regularly during their employment as well. Suppose that a drug test for an illegal drugs is 97% accurate in the case of a user of that drug, and 92% accurate in the case of a non-user for that drug. Suppose also that 5% of the entire population uses that drug.

- You are the hiring manager at a company that drug tests their employees. You have recently decided to hire a new employee. What is the prior probability that this employee is a user of this drug? (You may assume that this prospective employee is a randomly drawn person from the population.)
  - *P(drug user) = 0.05*

- The prospective employee gets drug tested, and the test comes out to be positive. What is the probability that they are actually a user for the drug? What is this probability called? Sketch a probability tree for this question.

- The prospective employee gets drug tested, and the test comes out to be positive. What is the probability that they are actually a user for the drug? What is this probability called? Sketch a probability tree for this question.
  - *P(drug user | +) $\rightarrow$ posterior probability*

- The prospective employee gets drug tested, and the test comes out to be positive. What is the probability that they are actually a user for the drug? What is this probability called? Sketch a probability tree for this question.
  - *P(drug user | +) $\rightarrow$ posterior probability*



- When the employee finds out that they tested positive, they refuse the test results, and say they would like to be tested again. What is the new prior probability you should use for this employee being a user of this drug?

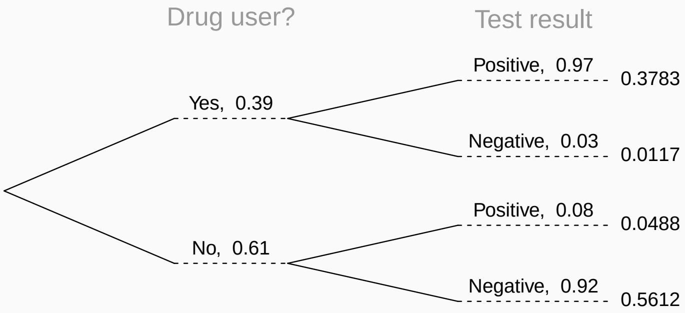
- When the employee finds out that they tested positive, they refuse the test results, and say they would like to be tested again. What is the new prior probability you should use for this employee being a user of this drug?
  - *P(drug user) = 0.39*

- When the employee finds out that they tested positive, they refuse the test results, and say they would like to be tested again. What is the new prior probability you should use for this employee being a user of this drug?
  - *P(drug user) = 0.39*
- The employee tests positive again in the second test. Should the new probability of them actually being a user of this drug be higher or lower than what you calculated before, or the same? Answer this question before you actually complete the calculations.

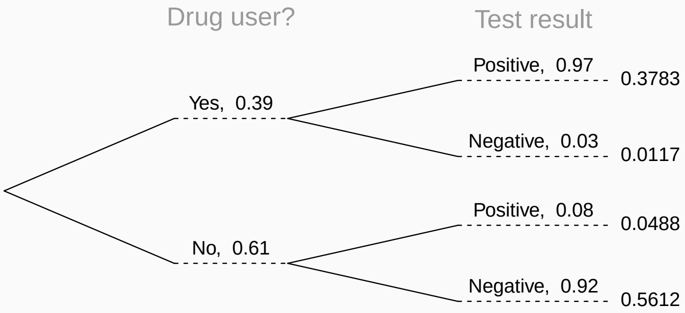
- When the employee finds out that they tested positive, they refuse the test results, and say they would like to be tested again. What is the new prior probability you should use for this employee being a user of this drug?
  - *P(drug user) = 0.39*
- The employee tests positive again in the second test. Should the new probability of them actually being a user of this drug be higher or lower than what you calculated before, or the same? Answer this question before you actually complete the calculations.
  - *Higher.*

- Finally, calculate the new updated probability that this employee is a user of this drug. When answering this question sketch a probability tree, take a picture, and upload to Sakai as an attachment.
  - *P(drug user | +) $\rightarrow$ posterior probability*

- Finally, calculate the new updated probability that this employee is a user of this drug. When answering this question sketch a probability tree, take a picture, and upload to Sakai as an attachment.
  - *P(drug user | +) → posterior probability*



- Finally, calculate the new updated probability that this employee is a user of this drug. When answering this question sketch a probability tree, take a picture, and upload to Sakai as an attachment.
  - *P(drug user | +) → posterior probability*



- Based on these results, would you hire this employee? Why or why not?
  - *No, quite likely that the employee is a drug user.*

Outline

3. Posterior probability and p-value do not mean the same thing

- *p-value* : P(observed or more extreme outcome | null hypothesis is true)
  - This is more like P(data | hyp) than P(hyp | data).
- *posterior* : P(hypothesis | data)
- Bayesian approach avoids the counter-intuitive Frequentist p-value for decision making, and more advanced Bayesian techniques offer flexibility not present in Frequentist models
- *Watch out!*
  - *Bayes*: A good prior helps, a bad prior hurts, but the prior matters less the more data you have.
  - *p-value*: It is really easy to mess up p-values: Goodman, 2008

Outline

Summary of main ideas

1. Probability trees are useful for conditional probability calculations
2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate
3. Posterior probability and p-value do not mean the same thing