Unit 2: Probability and distributions
2. Bayes' theorem and Bayesian inference

Sta 101 – Spring 2019

Duke University, Department of Statistical Science

Dr. Ellison

Slides posted at
https://www2.stat.duke.edu/courses/Spring19/sta101.001/

Outline

1. Housekeeping

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

3. Summary

Announcements

Coming up…

- Lab Assignment 2 is due **Thursday just before your lab section time.**
- Problem Set 2 due next **Wednesday 2/6 11:55**
- Readiness Assessment 3 next **Wednesday 2/6 in class!**

1

What we know:

- $P(A|B)=0.6$
- $P(A|\bar{B})=0.5$
- $P(B)=0.3$

Question: Is this enough information to find $P(B|A)$?

What we know:

- $P(A|B)$=0.6
- $P(A|\overline{B})$=0.5
- $P(B)$=0.3

Question: Is this enough information to find $P(B|A)$?

"observation"

Outline

1. Housekeeping

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

3. 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)

2

What we know:

- $P(A|B)$=0.6
- $P(A|\overline{B})$=0.5
- $P(B)$=0.3

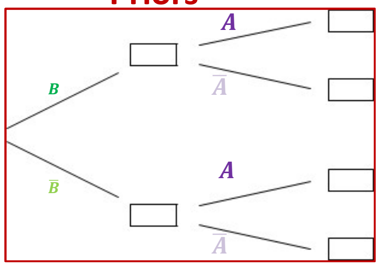
Question: Is this enough information to find $P(B|A)$?

Yes! We have enough information to fill out a Bayesian Probability Tree to get the answer!

What we know:
- $P(A|B)$=0.6
- $P(A|\overline{B})$=0.5
- $P(B)$=0.3

Want to know:
- $P(B|A)$

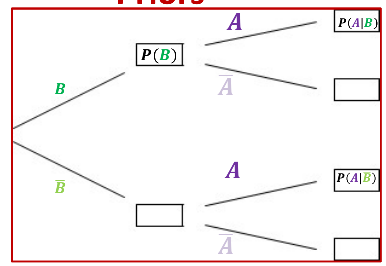
Priors



What we know:
- $P(A|B)$=0.6
- $P(A|\overline{B})$=0.5
- $P(B)$=0.3

Want to know:
- $P(B|A)$

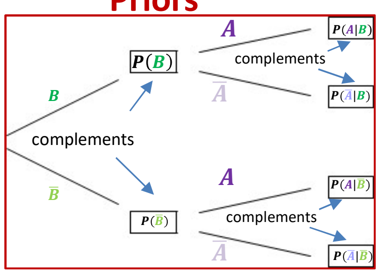
Priors



What we know:
- $P(A|B)$=0.6
- $P(A|\overline{B})$=0.5
- $P(B)$=0.3

Want to know:
- $P(B|A)$

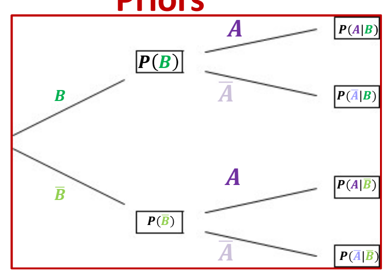
Priors



What we know:
- $P(A|B)$=0.6
- $P(A|\overline{B})$=0.5
- $P(B)$=0.3

Want to know:
- $P(B|A)$

Priors



Outline

1. Housekeeping

2. 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

3. Summary

Frequentist Inference

- Probability that makes decisions:

$$p-value = P\left(\begin{matrix}data\ more\ (or\ as)extreme\\ than\ observed\end{matrix} \middle| null\ hypothesis\ is\ true\right)$$

- All units except for 2.2
- More common
- Incorporate data all at once

Bayesian Inference

- Probability that makes decisions:

$$Posterior = P(hypothesis|observed\ data)$$

- Unit 2.2
- Less common
- Can incorporate data iteratively

Frequentist Inference

- Probability that makes decisions:

$$p-value = P\left(\begin{matrix}data\ more\ (or\ as)extreme\\ than\ observed\end{matrix} \middle| null\ hypothesis\ is\ true\right)$$

- All units except for 2.2
- More common
- Incorporate data all at once

Possible Outcomes

Fail to Reject Null Hypothesis | Reject Null Hypothesis

Bayesian Inference

- Probability that makes decisions:

$$Posterior = P(hypothesis|observed\ data)$$

- Unit 2.2
- Less common/newer
- Can incorporate data iteratively

Possible Outcomes

Range of Probabilities for a Hypothesis Being True

---

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**.

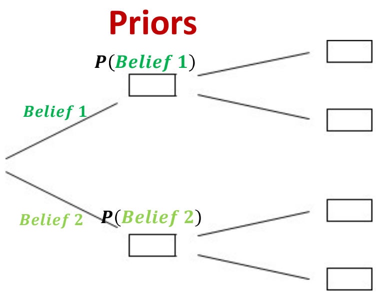
Want to know:

- $P(Belief\ 1|Data\ 1)$

3

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

- Step 1: You start with a set of **prior beliefs** (or prior probabilities).

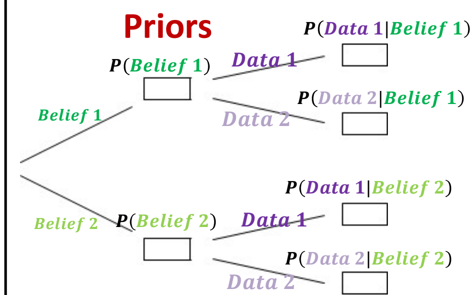


Want to know:
- $P(Belief\ 1|Data\ 1)$

3

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

- Step 1: You start with a set of **prior beliefs** (or prior probabilities).
- Step 2: You observe some data.
- Step 3: Based on that data, you update your beliefs.

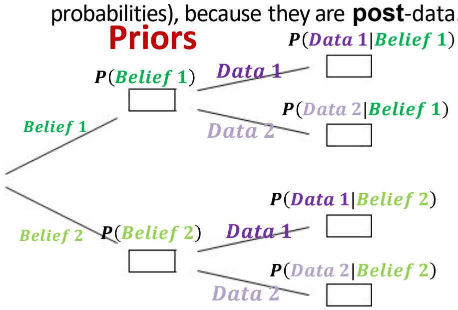


Want to know:
- $P(Belief\ 1|Data\ 1)$

3

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

- Step 1: You start with a set of **prior beliefs** (or prior probabilities).
- Step 2: You observe some data.
- Step 3: Based on that data, you update your beliefs.
- Step 4: These new beliefs are called **posterior beliefs** (or posterior probabilities), because they are **post**-data.



Want to know:
- $P(Belief\ 1|Data\ 1)$

3

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

Want to know:

Posterior Belief = $P(Belief\ 1|Data\ 1)$

3

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

Want to know:

Posterior Belief = $P(Belief\ 1|Data\ 1)$

$= \frac{P(Belief\ 1\ \text{and}\ Data\ 1)}{P(Data\ 1)}$

3

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

Want to know:

Posterior Belief = $P(Belief\ 1|Data\ 1)$

$= \frac{P(Belief\ 1\ \text{and}\ Data\ 1)}{P(Data\ 1)}$ Bayes Equation

3

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

Want to know:

Posterior Belief = $P(Belief\ 1|Data\ 1)$

$= \frac{P(Belief\ 1\ \text{and}\ Data\ 1)}{P(Data\ 1)}$

$= \frac{P(Data\ 1|Belief\ 1)P(Belief\ 1)}{P(Data\ 1)}$

3

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

Want to know:

Posterior Belief = $P(Belief\ 1|Data\ 1)$

$= \frac{P(Belief\ 1\ \text{and}\ Data\ 1)}{P(Data\ 1)}$

$= \frac{P(Data\ 1|Belief\ 1)P(Belief\ 1)}{P(Data\ 1)}$

General Multiplication Rule

3

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

Want to know:

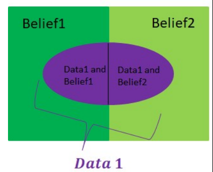
Posterior Belief = $P(Belief\ 1|Data\ 1)$

$$= \frac{P(Belief\ 1 \text{ and } Data\ 1)}{P(Data\ 1)}$$

$$= \frac{P(Data\ 1|Belief\ 1)P(Belief\ 1)}{P(Data\ 1)}$$

$$= \frac{P(Data\ 1|Belief\ 1)P(Belief\ 1)}{P(Data\ 1 \text{ and } Belief\ 1) + P(Data\ 1 \text{ and } Belief\ 2)}$$

*We assume that belief 1 and 2 cannot happen at the same time.



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

Want to know:

Posterior Belief = $P(Belief\ 1|Data\ 1)$

$$= \frac{P(Belief\ 1 \text{ and } Data\ 1)}{P(Data\ 1)}$$

$$= \frac{P(Data\ 1|Belief\ 1)P(Belief\ 1)}{P(Data\ 1)}$$

$$= \frac{P(Data\ 1|Belief\ 1)P(Belief\ 1)}{P(Data\ 1 \text{ and } Belief\ 1) + P(Data\ 1 \text{ and } Belief\ 2)}$$

$$= \frac{P(Data\ 1|Belief\ 1)P(Belief\ 1)}{P(Data\ 1|Belief\ 1)P(Belief\ 1) + P(Data\ 1|Belief\ 2)P(Belief\ 2)}$$

General Multiplication Rule

3

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

Final Equation we use with a Bayesian Probability Tree

Posterior Probability/Belief:

$$P(Belief\ 1|Data\ 1) = \frac{P(Data\ 1|Belief\ 1)P(Belief\ 1)}{P(Data\ 1|Belief\ 1)P(Belief\ 1) + P(Data\ 1|Belief\ 2)P(Belief\ 2)}$$

3

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**.
- Step 1: You start with a set of **prior beliefs** (or prior probabilities).
- Step 2: You observe some data.
- Step 3: Based on that data, you update your beliefs.
- Step 4: These new beliefs are called **posterior beliefs** (or posterior probabilities), because they are **post**-data.
- You can iterate this process.

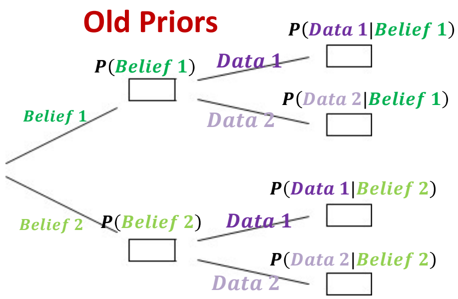
Step 5: Set **NEW prior beliefs = old posterior beliefs**

3

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

Iteration 1

- Observed Data 1
- **Found Posterior:** $P(Belief\ 1|Data\ 1)$



3

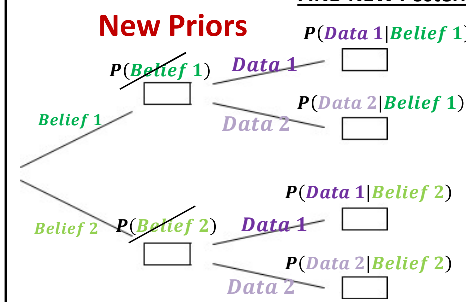
2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

Iteration 1

- Observed Data 1
- **Old Posterior:** $P(Belief\ 1|Data\ 1)$

Iteration 2

- NOW we observed Data 2
- **FIND NEW Posterior:** $P(Belief\ 1|Data\ 2)$



3

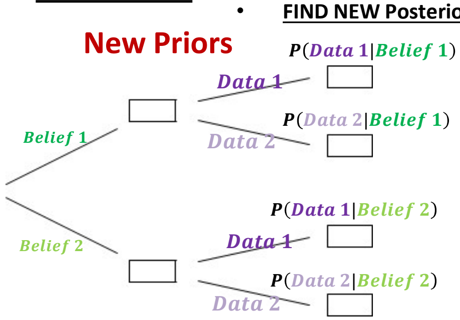
2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

Iteration 1

- Observed Data 1
- **Old Posterior:** $P(Belief\ 1|Data\ 1)$

Iteration 2

- NOW we observed Data 2
- **FIND NEW Posterior:** $P(Belief\ 1|Data\ 2)$



3

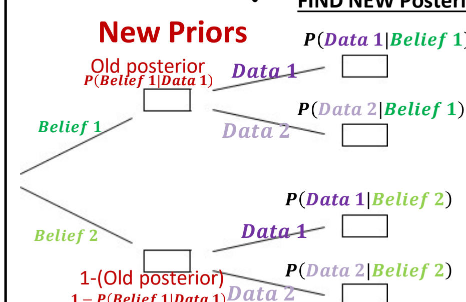
2. Bayesian inference: start with a prior, collect data, calculate posterior, make a decision or iterate

Iteration 1

- Observed Data 1
- **Old Posterior:** $P(Belief\ 1|Data\ 1)$

Iteration 2

- NOW we observed Data 2
- **FIND NEW Posterior:** $P(Belief\ 1|Data\ 2)$



3

Win candy today!

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

4

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
- The "good die" is the 12-sided die.

4

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
- 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"

4

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
- 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

4

Collecting Information on the Dice Location

Prior probabilities

- The two competing claims are
  - H_1: Good die is on left
  - H_2: Good die is on right

5

Prior probabilities

- The two competing claims are
  - H_1: Good die is on left
  - H_2: Good die is on right
- Winning in a Given Roll: The die you selected rolls a number ≥ 4
- At each roll I tell you whether you won or not (win = $\geq$ 4)
  - P(win | 6-sided die) =
  - P(win | 12-sided die) =

Question: If we knew which die was in which hand, which die would we always want to pick?

5

Prior probabilities

- The two competing claims are
  - H_1: Good die is on left
  - H_2: Good die is on right
- Winning in a Given Roll: The die you selected rolls a number ≥ 4
- At each roll I tell you whether you won or not (win = $\geq$ 4)
  - P(win | 6-sided die) = 3/6 = 0.5 → "Bad Die"
  - P(win | 12-sided die) = 9/12 = 0.75 → "Good Die"

5

Instructions for Volunteer Students who Collect Data

Rules of the game

Instructions for Student Volunteer

- You won't know which die I'm holding in which hand, left (L) or right (R).

In a Round for a Student Volunteer

- You pick die (L or R), I roll it, and I tell you if you win or not, where winning is getting a number $\geq$ 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.

6

Winning/Losing the Game

▶ Winning/Losing the GAME

Final Class Decision	*Truth* 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!*

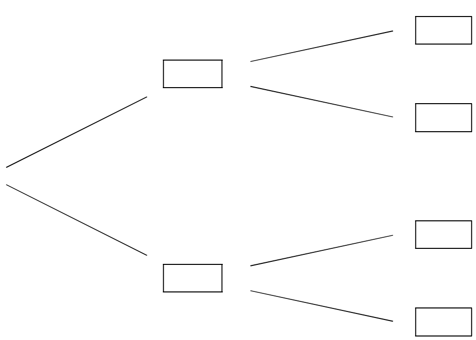
Sampling isn't free!

- 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?

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



1. Housekeeping

2. 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

3. Summary

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

11

Application exercise: 2.2 Bayesian inference for drug testing

See the course website for instructions.

12

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

13