UNIVERSITY of WASHINGTON

Data Science UW
Methods for Data Analysis

Conditional Probability
Lecture 4
Steve Elston

Presentation and story telling

Important part of data science

- Data science must have **impact**
- Results **only** have impact if they are understood
- Need to **'tell the story'**
- **Draw clear conclusion**
- Evidence supports conclusion

Presenting results is hard!

Tips for story telling

Make the story clear

- Occam's Razor
- You will only hold attention for a short time
- Don't distract your audience
- Start with your conclusion
- Support your conclusion with evidence
- Few words = greater impact!

Don’t obfuscate your message!

Short and simple has business impact

- Minimize discussion of methodology and technical detail
- Clear charts
  - Label axis
  - Minimize over-plotting
  - Simplify
- Short simple tables
  - Label rows and columns
  - Highlight key point
  - Minimal rows and columns

Don’t obfuscate your message!

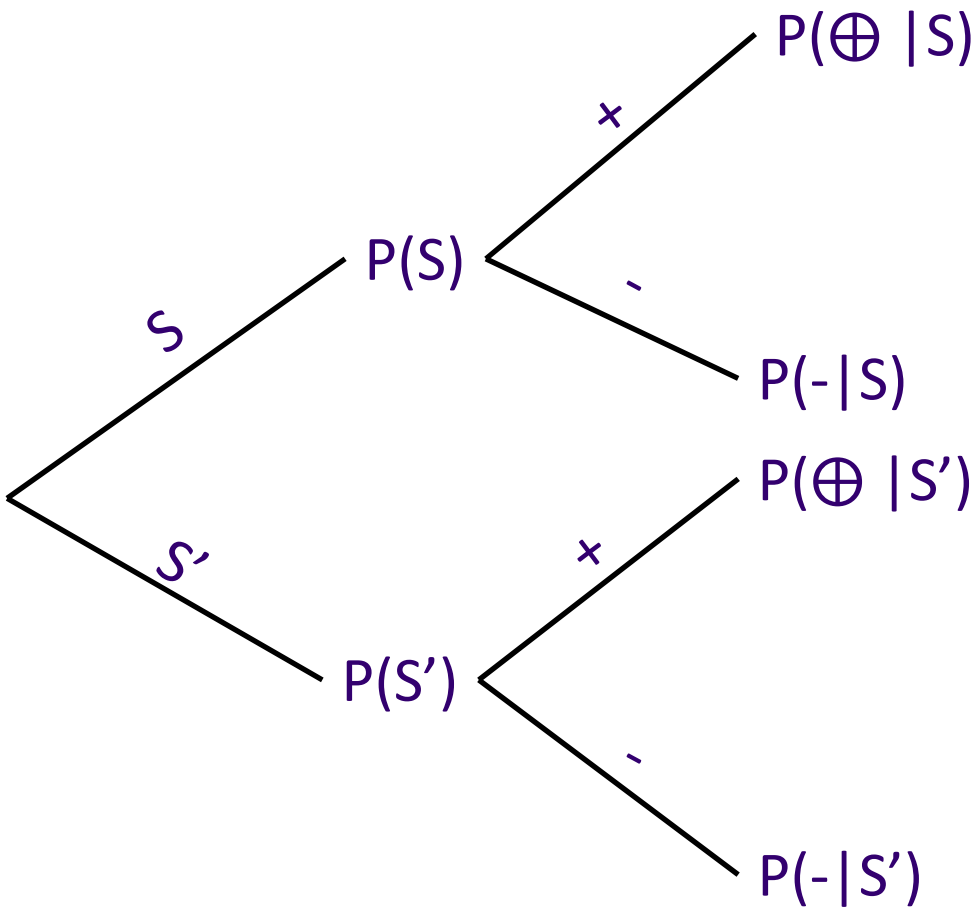
How to obfuscate your message!

- Include unnecessary detail
  - Long technical discussions
  - Discussion not focused on the problem
- Use obscure charts
  - No axis labels
  - Unneeded aesthetics
  - Two many or unneeded faceting
- Include unneeded tables
  - Use large complex tables
  - Don’t label rows and columns
  - Use charts that are unnecessary

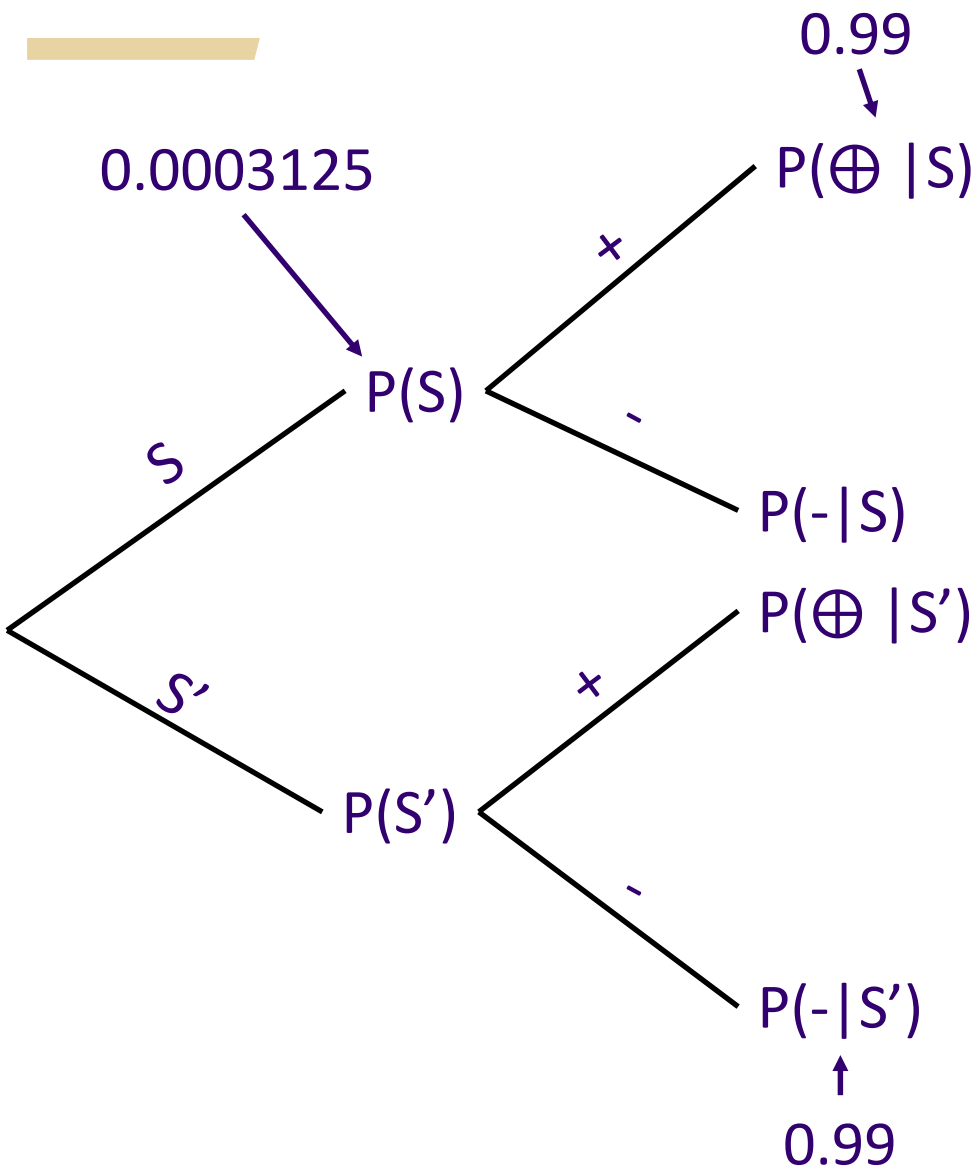
Conditional Probability Trees

- Let's consider a test for a Sickle Cell Anemia.
- Events:
  - S = patient has Sickle Cell Anemia
  - S' = patient does not have Sickle Cell Anemia
  - $\oplus$ = patient tests positive
  - - = patient tests negative
- Rate in US = 1/3200. P(S) = 1/3200 = 0.0003125.
- Medical company tells us that a test is 99% accurate.
  - $P(\oplus \mid S) = 0.99$
  - $P(- \mid S') = 0.99$

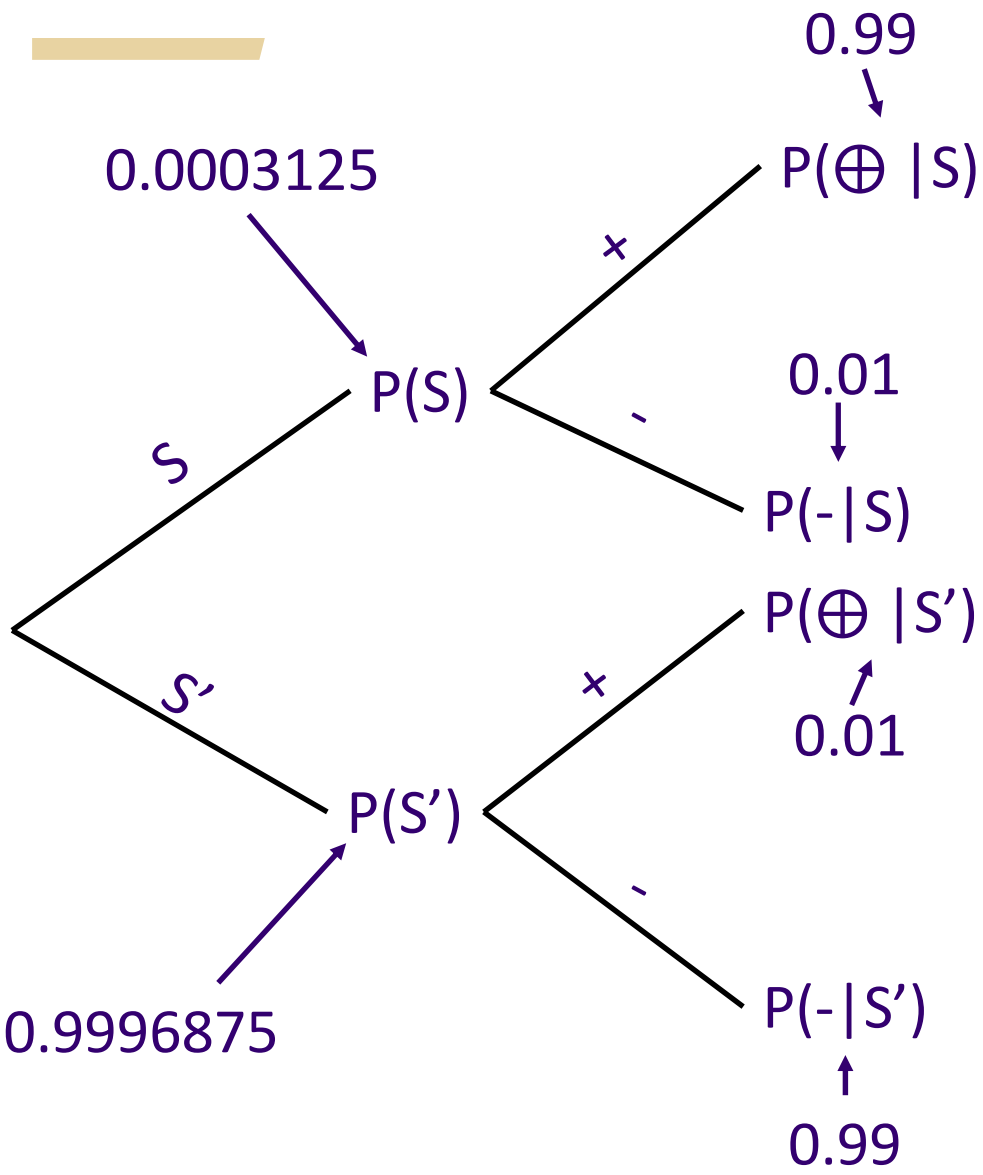
Conditional Probability Trees



Conditional Probability Trees



Conditional Probability Trees



Conditional Probability Trees

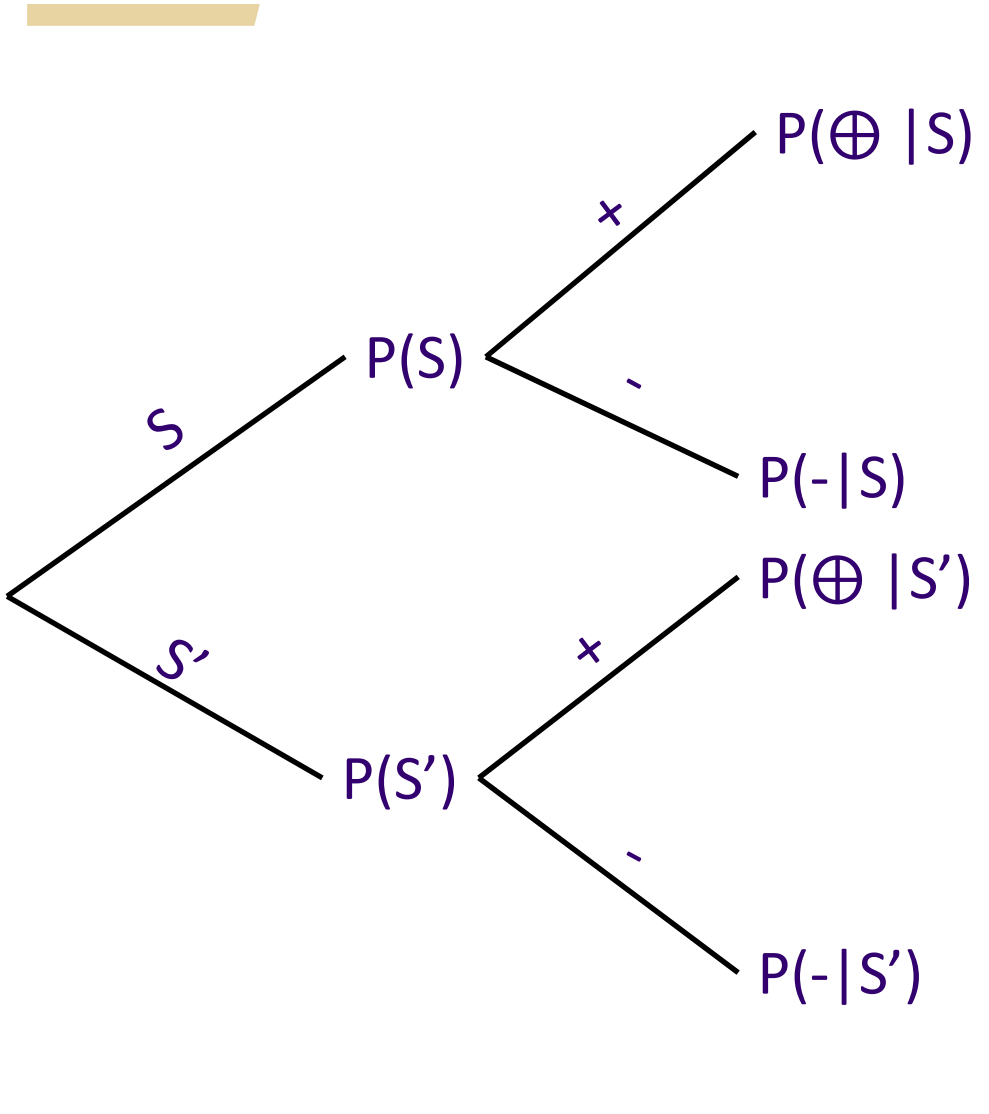
- What we really want to know is:
  - What is the P(S| ⊕)?
  - Also important to know: P(S|-)?
- From conditional probability definition:

$$P(S|\oplus) = \frac{P(S \cap \oplus)}{P(\oplus)}$$

- We also know that

$$P(\oplus) = P(\oplus \cap S) + P(\oplus \cap S')$$

Conditional Probability Trees



$P(\oplus \cap S) = P(S)P(\oplus|S)$

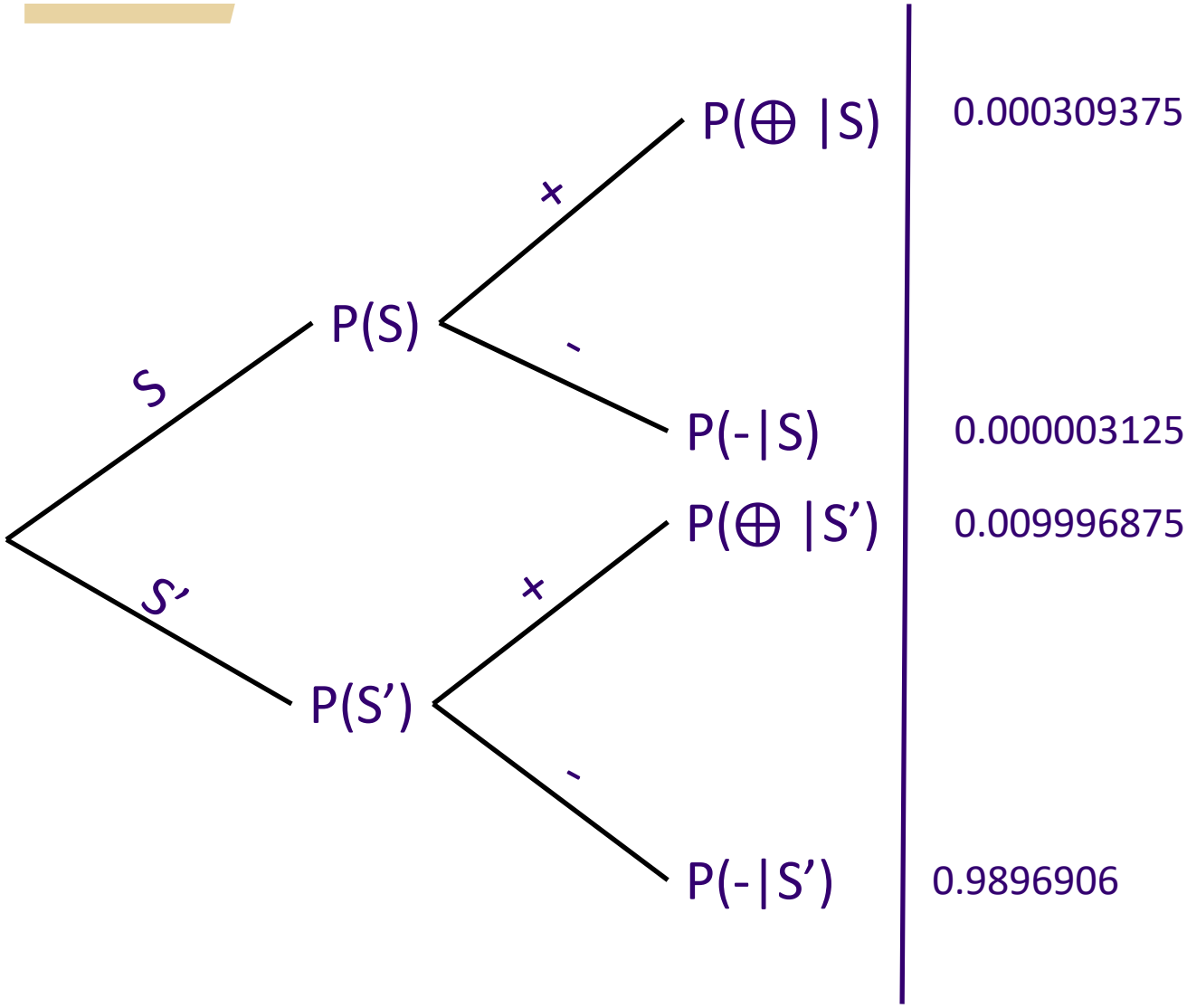
0.0003125*0.99=0.000309375

$P(- \cap S) = P(S)P(-|S)$

$P(\oplus \cap S') = P(S')P(\oplus|S')$

$P(- \cap S') = P(S)P(-|S')$

Conditional Probability Trees



Conditional Probability Trees

$$P(\oplus) = P(\oplus \cap S) + P(\oplus \cap S')$$

$$P(\oplus) = 0.01030625$$

0.000309375 = $P(\oplus \cap S)$

0.000003125 = $P(- \cap S)$

0.009996875 = $P(\oplus \cap S')$

0.9896906 = $P(- \cap S')$

$$P(S|\oplus) = \frac{P(S \cap \oplus)}{P(\oplus)}$$

$$P(S|\oplus) = \frac{0.000309375}{0.01030625}$$

$$\boldsymbol{P(S|\oplus) = 0.03001819}$$

Similarly,

$$P(S|-) = 0.000003157543$$

*Now see the probability interview question