The goal of this course is to introduce you to the fundamentals of probabilistic inference that underlie:

- Statistical data analysis
- Models of the acquisition and use of evidential knowledge regarding language structure

Probabilistic approaches are used increasingly in a wide variety of fields studying language.

I’ll give four simple examples of cases that we’ll have the tools to deal with by the end of the quarter.

We’ll spend the rest of today on the whiteboard going over some basics of probability.
Analyzing experimental data

- Does pragmatic knowledge affect rapid resolution of syntactic ambiguity?

  Mary *babysat* the children of the musician who...\
  
  - *are* generally arrogant and rude.
  - *is* generally arrogant and rude.

  Mary *detested* the children of the musician who...\
  
  - *are* generally arrogant and rude.
  - *is* generally arrogant and rude.

  *(Rohde, Levy, & Kehler, 2011)*

- One can carry out an experiment to study this (in our case, self-paced reading...*)
Analyzing experimental data

...and the raw measurements one gets out of the experiment look something like this:

How to infer whether the conditions are reliably different from one another?

(We hope that the blue and green curves are systematically shifted to the right more than the purple and salmon curves are)
Learning properties of a new dialect

• Imagine you’re a native English speaker thrust into a community speaking a dialect of English that’s novel for you
  • [e.g., Singaporean English]

• There may be some systematic distributional differences in the syntax of the languages
  • [e.g., passivization might be more or less frequent]

Sandy hired Jamie. vs. Jamie was hired by Sandy.

Somebody broke into the closet. vs. The closet was broken into.

(Wiener & Labov, 1984)
Learning properties of a new dialect

• The frequency of passivization is about 10% in American English (Rohde, Dick, & Elman, 2007)

• Suppose you hear 10 instances of passivizable sentences, 3 of which were in fact passivized

  Adam threw the ball.  
  Jane broke the vase.  
  Susan spotted Jerry.  
  Someone has eaten the cookies.  
  The manager called me up.  
  Someone cleaned the living room.  
  Don wrote this letter.  
  She was fired by the company.  
  This chair has been sat in.  
  This fact is already known by Pat.

• What proportion of the next several passivizable sentences you hear will be passivized?

  10%?  
  30%?  
  Somewhere in between?

• How confident should you be in your estimate of this chance?
Disentangling factors influencing speaker choice

- There are lots more cases where the grammar affords speakers a number of options to express the desired meaning.

  Terry gave the exhausted traveller from France a silver dollar.

  Terry gave a silver dollar to the exhausted traveller from France.

- Which do you prefer?

*(the dative alternation; Bresnan et al. 2007)*
Disentangling factors influencing speaker choice

• How about now?

Terry gave an exhausted traveller from France the silver dollar.

Terry gave the silver dollar to an exhausted traveller from France.

• [effects of definiteness]
Disentangling factors influencing speaker choice

• How about now?

  *Terry gave a traveller the silver dollar.*

  *Terry gave the silver dollar to a traveller.*

• [effects of phrase “weight”]

• In naturalistic language use, definiteness and weight are correlated! (*Wasow, 2002*)

• How do we disentangle the contributions of these factors to syntactic choice?
Cognitive modeling

- Phoneme perception from phonetic data
Cognitive modeling

- Identifying sequential dependencies in phoneme strings

\[ \text{da ta da ta da da da da da da ta ta ta da ta ta ta da da da da da} \]

- Random? Context-dependent?
Hierarchical models

• For both data analysis and cognitive modeling