Probability & statistics for linguists
Class 1: Motivation, basic probability

D. Lassiter
(h/t: R. Levy)
Why do (formal/psycho-/socio-/corpus-) linguists have need of a theory of uncertainty?

examples from your own research?
– data analysis (uncertainty about source of patterns)
– cognitive modeling (how ig. users cope with uncertainty)
why should linguists care? (1)

- language is variable, but highly structured grammatically and socially
  (Weinreich, Labov, & Herzog 1968)
- ‘knowledge of language’ – target for many theoretical linguists – includes much statistical knowledge, e.g.
  - sociolinguistics: Philly neighborhood studies
  - syntax: the dative alternation
why should linguists care? (2)

communication is a noisy-channel inference problem, with uncertainty at every step (Shannon, 1948)

languages can be thought of as coding schemes for human thought. are they efficient codes?
Why should linguists care? (3)

- **Corpus** and experimental methods are increasingly critical in linguistics.
- **Statistical analysis of data** is obligatory with these methods.
- Workflow: build prob. models, use to make predictions, test against data gathered.
- Deep theoretical understanding of probability is needed to do good statistical analysis.
why should linguists care? (4)

making good decisions despite uncertainty is the key to human intelligence

probability is indispensible in cog. science

– reasoning & decision-making
– perception
– motor control
– ...

relevant topics for many linguists; even for others, you’re human too...
What you’ll learn in this course

• fundamentals of probabilistic inference
• use in analysis of experimental data
• use in models of the acquisition and use of knowledge involving linguistic structure
  – smart inferences despite uncertainty, noise
  – coping with variability in the target structures [beyond ‘homogeneous community’ idealization]
• you’ll be able to deal with applications like the following [borrowed from R. Levy]
Analyzing experimental data

- Does pragmatic knowledge affect rapid resolution of syntactic ambiguity?

Mary *babysat* the children of the musician who...

Mary *detested* the children of the musician who...

\{ 
\begin{align*}
\text{are} & \text{ generally arrogant and rude.} \\
\text{is} & \text{ generally arrogant and rude.}
\end{align*}
\}

Analyzing experimental data

• 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?
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

  Sandy hired Jamie.

  VS.

  Jamie was hired by Sandy.

  Somebody broke into the closet.

  VS.

  The closet was broken into.

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

• 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.  
  Don wrote this letter.  
  She was fired by the company.  
  This chair has been sat in.  
  This fact is already known by Pat.  
  The manager called me up.  
  Someone cleaned the living room.

• 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?
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)_
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]
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?
phonemic perception

- Phoneme perception from phonetic data

![Graphs showing probability density and posterior probability of /b/ vs. VOT](image)
Hierarchical models

• For both data analysis and cognitive modeling
what is probability?

The theory of probabilities is nothing but good sense reduced to calculation; it allows one to appreciate with exactness what accurate minds feel by a sort of instinct ...  

-Laplace (1814)

Probability is not really about numbers; it is about the structure of reasoning.

-Shafer (1988)
what is probability?

probability is a logic

usually built on top of classical logic
  – an enrichment, not a competitor!

probability has a simple ‘possible worlds’ semantics
  probabilists’ “sample space” is just a partition of $W$, representing outcomes of interest
You and your friend meet at the park for a game of tennis. To decide who will serve first, you flip a coin. Your friend produces a quarter and tells you that it is a fair coin. What exactly does your friend mean by this?

- Frequentist: in many flips, proportion heads would approach .5
- Bayesian: he believes that heads and tails are equally likely outcomes
frequency and probability

You play ten matches, and your friend flips the same coin each time to determine who serves first. It lands heads six times and tails four.

• What are the **frequencies** of heads and tails in these flips?

• What are the **probabilities** of heads and tails for your friend's coin?

[unknown, but you have some idea]
inference

frequency ≠ probability

• We use frequencies of different outcomes to make **inferences** about probabilities.
• Frequentists and Bayesians differ in what procedures they think best for making inferences and for comparing models.
semantics of probability

sample space: $\Omega$

events $E$: all subsets of $\Omega$, including $\Omega$ itself.

Probability function:

1. for any event $E$, $P(E) \geq 0$.
2. $P(\Omega) = 1$.
3. for any $E_1$, $E_2$, if $E_1$ and $E_2$ are mutually exclusive, then $P(E_1 \text{ or } E_2) = P(E_1) + P(E_2)$.

(Kolmogorov, 1933)
example

hypothetical relative frequencies of the first word in an English sentence being each of the four open-class parts of speech:

N: 12.6%  
Adj: 2.1%  
V: 4.4%  
Adv: 8.9%

Suppose that these are the true probabilities, too. What can we derive about probabilities of various events in the sample space of possible first words of a sentence of English?
logistics

https://tr.im/a696d
http://web.stanford.edu/~danlass/courses/prob-and-stats-winter15/

or find thru “teaching” on my website

please monitor course website closely.
let’s take a look at the schedule (likely to change as quarter progresses)