**Course basics**

- Ling 681 Statistical Methods in Computational Linguistics
- Monday, Wednesday 7:00–8:15
- Instructor: Rob Malouf, rmalouf@mail.sdsu.edu
- Office hrs: BA 310A, Mondays 1:30-3:00, or by appointment
- http://rohan.sdsu.edu/~malouf/ling681.html

**Pre-/co-requisite:** Ling 581

**Lab**

**Python or perl?**

**Assessment:**
- Homeworks (20%)
- Take-home midterm (30%)
- Final project (50%)

**Requirements**

- **Textbook:**
  and

- **Additional readings**

**Schedule**

- **Week 1–4**  **Introduction** (Chapters 1, 2)
  *Background* · *Mathematical background* · *Probability* · *Information Theory*

- **Week 5–8**  **Statistics** (Chapters 3, 4, 5)
  *Descriptive statistics* · *Graphical methods* · *Hypothesis testing* · *Paired sign test* · *Bootstrap* · *Corpus statistics*

- **Week 9–10**  **Sequence models** (Chapters 6, 9, 10)
  *N grams* · *Smoothing* · *Hidden Markov models* · *Viterbi decoding* · *Part of speech tagging*

- **Week 11–14**  **Parsing** (Chapters 11, 12)
  *Probabilistic context free grammars* · *Inside–Outside algorithm* · *Treebank grammars* · *Dependency-based models*

- **Week 15**  **Class projects**
History

• Information theory, speech processing → statistical NLP
• Philology, formal language theory → symbolic NLP
• Competing paradigms
• Synthesis

Why probabilities?

• Continuous phenomena (speech)
• Machine learning
• Disambiguation
• Reasoning under uncertainty
• True randomness?

Statistics for linguists

• Basic statistical literacy is (or ought to be) a prerequisite for being a linguist in the 21st century.
• Statistical methods are an established part of 'peripheral' areas of linguistics (e.g., psycholinguistics, sociolinguistics, computational linguistics).
• Use of statistics in data collection and analysis is an essential part of the methodology of (virtually) all empirical sciences.
• The domain of linguistic theory is expanding to include inherently quantitative phenomena.

Adjective order

• From an idealized viewpoint, every sentence is either grammatical or it isn’t, and grammaticality is determined solely by the rules of the grammar.
• But:
  the large red American car
  *car American red the large
• Adjective order is incredibly complex, influenced by semantic, pragmatic, morphological, and collocational factors.
• But, it is relatively simple to build a statistical model that predicts adjective order with ~ 95% accuracy!
• But so?
Probability

- Roots in divination and gambling
- Rules of probabilities can be very unintuitive
- No scientific understanding of probability until 16th century
- Many paradoxes:
  - People use probability all the time for making decisions with partial information, yet seem to be hardwired to believe everything follows deterministic rules
  - Random selection is used to make ‘fair’ decisions, yet can lead to very unfair outcomes

Wanna bet?

- Someone is willing to pay even money ($1 to your $1) if you can roll a four or better. Is this a good bet?

- A correct method for answering questions like this was first given in Gerolamo Cardano’s Liber de Ludo Aleae, a gambler’s manual written in 1569 but not published until 1663.

Wanna bet?

- Cardano’s method is to compare the number of favorable outcomes with the number of unfavorable outcomes.

- Possible outcomes:

<table>
<thead>
<tr>
<th>Lose</th>
<th>Win</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

- The odds of winning are 3 : 3 or 1 : 1

Wanna bet?

- Now someone offers you $2 to your $1 if you can get at least one six on two throws of a die. Is this a good bet?

- To use Cardano’s method, we need to know how many possible outcomes there are.

- There are six outcomes for one throw, and each of those could followed up by six possible second throws: $6 \times 6 = 36$

- How many times to we win? There are six ways to win on the first throw, and five ways to win on the second: $5 + 6 = 11$

- So, the odds are 25 : 11 against us, or about 2.3 : 1

- Prior to Cardano, gambler’s folklore put the odds at 24 : 12 or 2 : 1
Probability

- Gamblers tend to talk about chance outcomes in terms of odds, the ratio of favorable outcomes to unfavorable outcomes.

- In other contexts, it is usually easier to talk about probability, the ratio of favorable outcomes to total possible outcomes:
  \[ p = \frac{f}{n} \]

- For impossibilities, there are no favorable outcomes:
  \[ p = \frac{0}{n} = 0 \]

- For certainties, all outcomes are favorable:
  \[ p = \frac{n}{n} = 1 \]

Random experiments

- To make reasoning about more complicated games easier, we want a mathematical model which abstracts away from some the details.

- A random experiment or trial is a observation where:
  - All possible outcomes are known in advance;
  - The actual outcome is not known in advance; and
  - The experiment can be carried out repeatedly under identical conditions.

- For a random experiment, the set of possible outcomes is the sample space \( S \)

Random variables

- A random variable \( X \) is a function from a sample space to the set of real numbers:
  \[ X : S \to \mathbb{R} \]

- For example, if the experiment is one roll of a die, the sample space is:
  \[ S = \{\text{1, 2, 3, 4, 5, 6}\} \]

- For this sample set, a random variable \( X \) might be the number of pips showing. Another random variable \( Y \) might be the number of 6's thrown.

Events

- Suppose \( X \) is a random variable for an experiment with sample space \( S \). An event is a statement about \( X \) that picks out a subset of \( S \):
  \[ X \in B = \{s \in S : X(s) \in B\} \]

- For example, if the random variable \( X \) is the number of pips on one roll, then one interesting event is \( X \geq 4 \), the set of outcomes:
  \[ \{s \in S : X(s) \geq 4\} \]

- For any \( A \subset S \), we can assume an indicator variable \( A \)

- The set of all possible events, the event space \( \mathcal{F} \), is the power set of \( S \) (i.e., the set of all possible subsets of the sample space).
Probability

A probability distribution is a function from an event space to a real number between 0 and 1 (inclusive):

\[ P : \mathcal{F} \rightarrow [0, 1] \]

that satisfies the conditions:

- The probability of a certain event is one:
  \[ P(S) = 1 \]

- For any mutually exclusive events \( A \) and \( B \):
  \[ P(A \cup B) = P(A) + P(B) \]