Lectures 2-6: The mathematical stuff

Statistical Methods for Natural Language Processing
Fredrik Engström

February 3, 2011

Presentation

Fredrik Engström
email: fredrik.engstrom@gu.se
phone: 031 – 786 63 35

- Lecture 2: Basic probability theory
- Lecture 3: Bayesian updating + Random variables
- Lecture 4: Random variabels + Information theory
- Exercise session 1:
- Lecture 5: Statistical inference
- Lecture 6: (Hidden) Markov chains
- Exercise session 2:

Preliminary example: The hit-and-run

- Two taxi companies, one with green and one with blue cars.
- 85% of all the taxi cars are blue.
- A person witnesses a hit-and-run accident involving a taxi car.
- The witness observes a green car.
- Turns out that the witness has a reliability of 80% of determining the right color of a car.

What is the probability that the car was green?

Set theory

See `http://en.wikipedia.org/wiki/Set_(mathematics)`.

- A **set** is a collection of objects (abstract or concrete).
- Examples: $\{5, 32, 9\}$, $\{\texttt{hej}, \texttt{vem}, \texttt{jag}\}$.
- $a \in A$
- Unordered. Without repetitions. $\{5, 32, 9\} = \{32, 9, 32, 5, 9\}$
- **Subset**
- **Union**
- **Intersection**
- **Emptyset**
- **Set difference** / **Complement**
- Singletons not equal to its element.
- **disjoint** union and **partition**

Probabilities properly

- Ω is the **sample set**. (Set of outcomes.)
- **Events** are subsets of Ω.
- A **probability function** P assigns probabilities to events.
- $P(A)$ is the probability of the event A.
- P has to be such that:
  - $0 \leq P(A) \leq 1$ for every event A.
  - $P(\Omega) = 1$
  - $P(A \cup B) = P(A) + P(B)$ if A and B are disjoint, i.e., $A \cap B = \emptyset$.

  (The axioms of a probability function.)

Examples: Two coin tosses. A man with two children.

Conditional probabilities

Definition

If $P(B) \neq 0$ then

$$\mathbf{P(A|B)} = \frac{P(A \cap B)}{P(B)}.$$

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$$

Theorem (Bayes theorem)

If $P(A) \neq 0$ *then*

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

Dependence / Independence

Intuition: Two events A and B are independent if the probability of A does not depend on whatever assuming B or not:

$$P(A|B) = P(A).$$

Definition

A and B are **independent** if $P(A \cap B) = P(A)P(B)$.

(This definition works even for $P(A) = P(B) = 0$.)
Example: pos-tagged words. Rolling a die.

Summary

- P satisfies:
  - $0 \leq P(A) \leq 1$,
  - $P(\Omega) = 1$,
  - $P(A \cup B) = P(A) + P(B)$ whenever $A \cap B = \emptyset$.
- $P(A|B) = P(A \cap B)/P(B)$
- $P(A|B) = P(B|A)P(A)/P(B)$.
- A and B are independent if $P(A \cap B) = P(A)P(B)$.