

**CA446: Statistical Machine Translation**  
**Exercises**  
**Thursday, 20th February 2014**

1. Part-of-speech tag the following sentences:

- (a) Time flies like an arrow.
- (b) Joe reads the book by the famous author to his sister.
- (c) It is getting wet because it is raining.
- (d) Jane likes to go to Hoboken.

Use the following part-of-speech tags: NOUN, VERB, ADJ, DET, ADV, PRON, CONJ, NUM, PREP, INF

2. If we flip a coin 10 times, we might get the outcome *HTTHTHTHTT* (*H* for *heads*, *T* for *tails*).

- (a) Estimate a distribution by maximum likelihood estimation.
- (b) We want to test the quality of the estimation. We flip the coin five times and get *HHTTH*. What is the probability of this outcome according to (1) the estimated distribution, and (2) the uniform distribution?

3. (a) Prove that  $p(y|x) = p(y)$  if  $X$  and  $Y$  are independent.

(b) Derive Bayes theorem.

4. Given that the main uses of translation technology are assimilation, dissemination, and communication, describe how these uses differ in terms of

- (a) demands for quality
- (b) requirements for speed
- (c) willingness to spend money
- (d) size of the market