CS 585 Natural Language Processing
Fall 2004
Homework 2: Probabilities and Information Theory

Out: Thu, September 23, 2004
Due: Thu, October 5, 2004

1. You roll a die, then roll it again. What is the probability that you get the same number from both rolls? Explain in terms of event spaces and basic outcomes.

2. Make up and write down a conditional probability table for $P(W|O, P)$, where $W \in$ { democrats, republicans, green party } win the Fall election, and $O \in$ {bin Laden is captured, he is not captured }, and $P \in$ {Los Angeles has to be evacuated due to polution problems, it doesn't}.

3. Suppose one is interested in a rare syntatic construction, parasitic gaps, which occur on average once in 100,000 sentences. Peggy Linguist has developped a complicated pattern matcher that attempts to identify sentences with parasitic gaps. It's pretty good, but its not perfect: if a sentence has a parastic gap, it will say so with probability 0.95, if it doesn't it will wrongly say so with probability 0.005. Suppose the test says that a sentence contains a parasitic gap. What is the probability that this is true?

4. I have a fair 4-sided die that is red. I have a fair 8-sided die that is blue. Let X be a random variable over numbers 1 through 8. Let C be a uniformly-distributed random variable over the color die I roll, red or blue. Recall that entropy of a random variable, Y, is $H(Y) = \sum_i p(y_i) \log_2 p(y_i)$. What is the entropy of X given that I use the blue die, $H(X|C = \text{blue})$? What is the mutual information between X and C, $I(X; C)$?

5. Given a corpus consisting of *aabccccc*, what are the maximum likelihood estimates for $P(X = a), P(X = b), P(X = c), P(X = d)$?

6. If $P(X = a) = 0.0$, $P(X = b) = 0.25$, $P(X = c) = 0.25$ and $P(X = d) = 0.5$, what is the entropy of random variable X? What is the cross entropy of X with a corpus consisting of *ababc*? What distribution would have lower cross-entropy?

7. **Extra credit**: You have a biased k-sided die, with probabilities for each face $q_1, q_2, ..q_i, ...q_k$. You roll it n times and obtain counts for each face, $m_1, m_2, ...m_i, ...m_k$. The maximum likelihood estimate of q_i is the classic and intuitive ratio of counts, $\frac{m_i}{n}$. Prove this. (Similar to our in-class proof for the binomial maximum likelihood.)