BMI/CS 776 Spring 2008
Homework #6

Prof. Colin Dewey

Due Tuesday, May 6th, 2008 by 11:59pm

The goal of this assignment is to become familiar with *stochastic context-free grammars (SCFGs)*. You have three options for turning in this homework:

- Copy all relevant files to the directory:
  `/u/medinfo/handin/bmi776/hw6/USERNAME`
  where `USERNAME` is your account name for the BMI network.
- Send it to me by email
- Turn it in on paper during class on Tuesday, May 6th or put it in my mailbox by 5pm on that day.

1. (Textbook exercise 10.3) As we have given it, the Nussinov algorithm can produce nonsensical 'base pairs' between adjacent complementary residues. Modify the Nussinov folding algorithm so that hairpin loops must have a minimum length of h. Give the modifications that are required for the fill and traceback steps.

2. (Textbook exercise 9.6) Consider the complete language generated by the CFG:

$$
\begin{aligned}
S &\rightarrow aW_1u \mid cW_1g \mid gW_1c \mid uW_1a \\
W_1 &\rightarrow aW_2u \mid cW_2g \mid gW_2c \mid uW_2a \\
W_2 &\rightarrow aW_3u \mid cW_3g \mid gW_3c \mid uW_3a \\
W_3 &\rightarrow gaaa \mid gcaa
\end{aligned}
$$

   Describe a regular grammar that generates *exactly* the same language. Does describing this sequence family with a regular grammar seem like a good idea?

3. Consider the following RNA (only consisting of `a` and `u`) SCFG:

$$
\begin{array}{llllllllllll}
S & \rightarrow & T, & S & \rightarrow & L & & & & & & \\
 & 0.8 & & & 0.2 & & & & & & & \\
T & \rightarrow & \mathrm{a}T\mathrm{u} & T & \rightarrow & \mathrm{u}T\mathrm{a} & T & \rightarrow & \mathrm{a}T & T & \rightarrow & L \\
 & 0.3 & & & 0.3 & & & 0.3 & & & 0.1 & \\
L & \rightarrow & \mathrm{a}L & L & \rightarrow & \mathrm{u}L & L & \rightarrow & \mathrm{a} & L & \rightarrow & \mathrm{u} \\
 & 0.4 & & & 0.4 & & & 0.1 & & & 0.1 &
\end{array}
$$

(a) Give one derivation for the string `aau` using this grammar. What is the (prior) probability of this derivation?

(b) What is the likelihood of the string `aau` given the grammar?

(c) What is the posterior probability of the derivation from (a)?