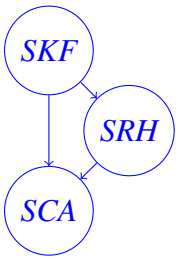
Ling 5801: Problem Set 5

Due via Carmen dropbox at 11:59 PM 11/14.

1. [10 pts.] Write an equation for a full joint distribution in terms of the following models:

   - θ_{SKF}, for a student knowing the fact that answers a question,
   - θ_{SRH}, for the student raising his or her hand,
   - θ_{SCA}, for the student correctly answering the question,

   with conditional dependencies as shown in the following network:



2. [10 pts.] Draw or describe a graphical representation of an extension of the above probability model, using random variables for:

   - student listening to lesson explaining fact
   - teacher asking question
   - student hearing question

   Justify each additional conditional dependency in a sentence (for example: 'a student is more likely to raise his or her hand if he or she knows the answer').

3. [10 pts.] PROGRAMMING: Write a program to read in models of language change over generations of speakers. Use the following format for component models of a grandparent speaker G, a parent speaker P (given grandparent), and a child speaker C (given parent) making use of the word 'who' as opposed to 'whom' in the position of an accusative filler (e.g. 'who/whom did you invite?'):

```
G : who  = .1
G : whom = .9

⋮

P who  : who  = 1
P who  : whom = 0
P whom : who  = .2
P whom : whom = .8

⋮
```

```
C who  : who  = 1
C who  : whom = 0
C whom : who  = .5
C whom : whom = .5
```

⋮

then use these models to calculate a conditional probability distribution table for $\mathsf{P}(P \mid C)$, and print it in the following format:

```
PgivC who : who = 0.4375
PgivC who : whom = 0.5625
```

⋮

4. [10 pts.] PROGRAMMING: Write a program to read in models for all variables R, W, and O in the 'repeated trials' model shown at the beginning of the lecture notes on sequence modeling, in the following format

```
R : ohio = .5
R : phil = .5
W : /nek/ = .6
W : /naek/ = .4
O ohio /nek/ : [nek] = 1
O phil /nek/ : [nek] = .667
O phil /nek/ : [naek] = .333
O ohio /naek/ : [naek] = 1
O phil /naek/ : [naek] = 1
```

and an input sequence of any number of observations in the format:

```
I [naek] [nek] [naek] ...
```

then print out a probability distribution for R given all of these input observations, in the following format (note: probabilities given observations should not necessarily match initial R model):

```
RgivenIdata : ohio = .4
RgivenIdata : phil = .6
```

Your program should be as short as possible. Hand in all inputs and outputs.