**Out:** *Tue Oct 9*
Due: *Tue Oct 16 (beginning of class)*
Supplementary reading: RN, Ch 14.1-14.3

2.1 Probabilistic reasoning

A patient is known to have contracted a rare disease which comes in two forms, represented by the values of a binary random variable $D \in \{0, 1\}$. Symptoms of the disease are represented by the binary random variables $S_k \in \{0, 1\}$, and knowledge of the disease is summarized by the belief network:
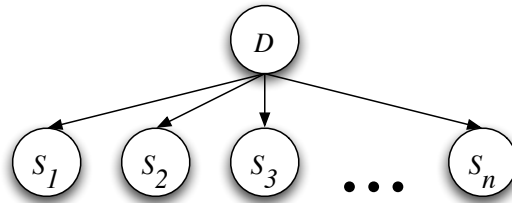
The conditional probability tables (CPTs) for this belief network are as follows. In the absence of evidence, both forms of the disease are equally likely, with prior probabilities: $P(D=0) = P(D=1) = \frac{1}{2}$. In one form of the disease $(D=0)$, all the symptoms are uniformly likely to be observed, with $P(S_k=0|D=0) = \frac{1}{2}$ for all k. By contrast, in the other form of the disease $(D=1)$, the first symptom occurs with probability one,

$$P(S_1=1|D=1) = 1,$$

while the kth symptom (with $k \geq 2$) occurs with probability

$$P(S_k=1|D=1) = \frac{f(k-1)}{f(k)},$$

where the function $f(k)$ is defined by

$$f(k) = 2^k + (-1)^k.$$

Suppose that on the kth day of the month, a test is done to determine whether the patient is exhibiting the kth symptom, and that each such test returns a positive result. Thus, on the kth day, the doctor observes the patient with symptoms $\{S_1=1, S_2=1, \ldots, S_k=1\}$. Based on the cumulative evidence, the doctor makes a new diagnosis each day by computing the ratio:

$$r_k = \frac{P(D=0|S_1=1, S_2=1, \ldots, S_k=1)}{P(D=1|S_1=1, S_2=1, \ldots, S_k=1)}.$$

If this ratio is greater than 1, the doctor diagnoses the patient with the $D=0$ form of the disease; otherwise, with the $D=1$ form.

(a) Compute the ratio r_k as a function of k. How does the doctor's diagnosis depend on the day of the month? Show your work.

(b) Does the diagnosis become more or less certain as more symptoms are observed? Explain.

2.2 Noisy-OR
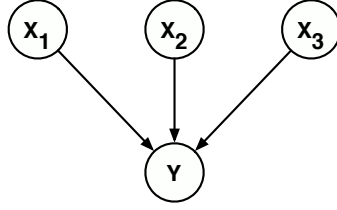
(a) **Conditional probability table**

Complete the noisy-OR conditional probability table for the belief network shown above. (The missing values in the table are determined by the ones shown.)

X_1	X_2	X_3	$P(Y = 1 \mid X_1, X_2, X_3)$
0	0	0	0
1	0	0	
0	1	0	$\frac{1}{3}$
0	0	1	
1	1	0	
1	0	1	
0	1	1	$\frac{4}{5}$
1	1	1	$\frac{5}{6}$

(b) **Qualitative reasoning**

Suppose that each node X_i in this model has some finite, non-zero prior probability to be either zero or one; namely $0 < P(X_i{=}1) < 1$. Sort the following probabilities from smallest to largest:

$$P(X_2{=}1)$$
$$P(X_2{=}1 \mid Y{=}0)$$
$$P(X_2{=}1 \mid Y{=}1)$$
$$P(X_2{=}1 \mid Y{=}1, X_1{=}0, X_3{=}0)$$
$$P(X_2{=}1 \mid Y{=}1, X_1{=}1, X_3{=}1)$$

For this part, you are **not** required to compute numerical values for these probabilities, only to sort them in ascending order. In fact, this part can be completed independently of the values in part (a).

(c) **Quantitative reasoning**

Suppose that $P(X_i = 1) = \frac{1}{3}$ for $i \in \{1, 2, 3\}$. Compute numerical values for the probabilities in part (b) from these priors and your answers in part (a). Do these values match your previous ordering?

2.3 Hangman

Consider the belief network shown below, where the random variable W stores a five-letter word and the random variable $L_i \in \{\texttt{A}, \texttt{B}, \ldots, \texttt{Z}\}$ reveals only the word's ith letter. Also, suppose that these five-letter words are chosen at random from a large corpus of text according to their frequency:

$$P(W\!=\!w) \;=\; \frac{\text{COUNT}(w)}{\sum_{w'} \text{COUNT}(w')},$$

where $\text{COUNT}(w)$ denotes the number of times that w appears in the corpus and where the denominator is a sum over all five-letter words. Note that in this model the conditional probability tables for the random variables L_i are particularly simple:

$$P(L_i\!=\!\ell|W\!=\!w) \;=\; \begin{cases} 1 & \text{if } \ell \text{ is the } i\text{th letter of } w, \\ 0 & \text{otherwise.} \end{cases}$$

Now imagine a game in which you are asked to guess the word w one letter at a time. The rules of this game are as follows: after each letter (\texttt{A} through \texttt{Z}) that you guess, you'll be told whether the letter appears in the word and also where it appears. Given the *evidence* that you have at any stage in this game, the critical question is what letter to guess next.
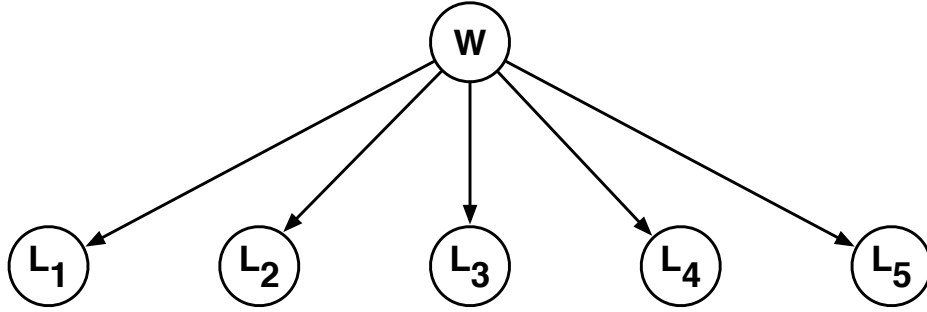
Let's work an example. Suppose that after three guesses—the letters \texttt{D}, \texttt{I}, \texttt{M}—you've learned that the letter \texttt{I} does *not* appear, and that the letters \texttt{D} and \texttt{M} appear as follows:

$$\underline{\;\;\texttt{M}\;\;} \quad \underline{\;\;\;\;\;} \quad \underline{\;\;\texttt{D}\;\;} \quad \underline{\;\;\;\;\;} \quad \underline{\;\;\texttt{M}\;\;}$$

Now consider your next guess: call it ℓ. In this game the best guess is the letter ℓ that maximizes

$$P\Big(L_2\!=\!\ell \textbf{ or } L_4\!=\!\ell \;\Big|\; L_1\!=\!\texttt{M}, L_3\!=\!\texttt{D}, L_5\!=\!\texttt{M}, L_2 \notin \{\texttt{D}, \texttt{I}, \texttt{M}\}, L_4 \notin \{\texttt{D}, \texttt{I}, \texttt{M}\}\Big).$$

In other words, pick the letter ℓ that is most likely to appear in the blank (unguessed) spaces of the word. For any letter ℓ we can compute this probability as follows:

$$P\Big(L_2\!=\!\ell \textbf{ or } L_4\!=\!\ell \;\Big|\; L_1\!=\!\texttt{M}, L_3\!=\!\texttt{D}, L_5\!=\!\texttt{M}, L_2 \notin \{\texttt{D}, \texttt{I}, \texttt{M}\}, L_4 \notin \{\texttt{D}, \texttt{I}, \texttt{M}\}\Big)$$

$$= \sum_w P\Big(W\!=\!w, L_2\!=\!\ell \textbf{ or } L_4\!=\!\ell \;\Big|\; L_1\!=\!\texttt{M}, L_3\!=\!\texttt{D}, L_5\!=\!\texttt{M}, L_2 \notin \{\texttt{D}, \texttt{I}, \texttt{M}\}, L_4 \notin \{\texttt{D}, \texttt{I}, \texttt{M}\}\Big), \quad \boxed{\textbf{marginalization}}$$

$$= \sum_w P\Big(W\!=\!w | L_1\!=\!\texttt{M}, L_3\!=\!\texttt{D}, L_5\!=\!\texttt{M}, L_2 \notin \{\texttt{D}, \texttt{I}, \texttt{M}\}, L_4 \notin \{\texttt{D}, \texttt{I}, \texttt{M}\}\Big)\, P\big(L_2\!=\!\ell \textbf{ or } L_4\!=\!\ell | W\!=\!w\big) \boxed{\textbf{product rule \& CI}}$$

where in the third line we have exploited the conditional independence (**CI**) of the letters L_i given the word W. Inside this sum there are two terms, and they are both easy to compute. In particular, the second term is more or less trivial:

$$P(L_2=\ell \text{ or } L_4=\ell|W=w) = \begin{cases} 1 & \text{if } \ell \text{ is the second or fourth letter of } w \\ 0 & \text{otherwise.} \end{cases}$$

And the first term we obtain from Bayes' rule:

$$P(W=w|L_1=\texttt{M}, L_3=\texttt{D}, L_5=\texttt{M}, L_2 \notin \{\texttt{D},\texttt{I},\texttt{M}\}, L_4 \notin \{\texttt{D},\texttt{I},\texttt{M}\})$$

$$= \frac{P\big(L_1=\texttt{M}, L_3=\texttt{D}, L_5=\texttt{M}, L_2 \notin \{\texttt{D},\texttt{I},\texttt{M}\}, L_4 \notin \{\texttt{D},\texttt{I},\texttt{M}\}|W=w\big)P(W=w)}{P(L_1=\texttt{M}, L_3=\texttt{D}, L_5=\texttt{M}, L_2 \notin \{\texttt{D},\texttt{I},\texttt{M}\}, L_4 \notin \{\texttt{D},\texttt{I},\texttt{M}\})} \quad \boxed{\textbf{Bayes' rule}}$$

In the numerator of Bayes' rule are two terms; the left term is equal to zero or one (depending on whether the evidence is compatible with the word w), and the right term is the prior probability $P(W = w)$, as determined by the empirical word frequencies. The denominator of Bayes' rule is given by:

$$P(L_1=\texttt{M}, L_3=\texttt{D}, L_5=\texttt{M}, L_2 \notin \{\texttt{D},\texttt{I},\texttt{M}\}, L_4 \notin \{\texttt{D},\texttt{I},\texttt{M}\})$$
$$= \sum_w P(W=w, L_1=\texttt{M}, L_3=\texttt{D}, L_5=\texttt{M}, L_2 \notin \{\texttt{D},\texttt{I},\texttt{M}\}, L_4 \notin \{\texttt{D},\texttt{I},\texttt{M}\}), \quad \boxed{\textbf{marginalization}}$$
$$= \sum_w P(W=w)P(L_1=\texttt{M}, L_3=\texttt{D}, L_5=\texttt{M}, L_2 \notin \{\texttt{D},\texttt{I},\texttt{M}\}, L_4 \notin \{\texttt{D},\texttt{I},\texttt{M}\}|W=w), \quad \boxed{\textbf{product rule}}$$

where again all the right terms inside the sum are equal to zero or one. Note that the denominator merely sums the empirical frequencies of words that are compatible with the observed evidence.

Now let's consider the general problem. Let E denote the evidence at some intermediate round of the game: in general, some letters will have been guessed correctly and their places revealed in the word, while other letters will have been guessed incorrectly and thus revealed to be absent. There are two essential computations. The first is the *posterior* probability, obtained from Bayes' rule:

$$P(W=w|E) = \frac{P(E|W=w)\,P(W=w)}{\sum_{w'} P(E|W=w')\,P(W=w')}.$$

The second key computation is the *predictive* probability, based on the evidence, that the letter ℓ appears somewhere in the word:

$$P\big(L_i=\ell \text{ for some } i \in \{1,2,3,4,5\}\big|E\big) = \sum_w P\big(L_i=\ell \text{ for some } i \in \{1,2,3,4,5\}\big|W=w\big)P\big(W=w\big|E\big).$$

Note in particular how the first computation feeds into the second. Your assignment in this problem is to implement both of these calculations. **You may program in the language of your choice.**

(a) Download the file *hw2_word_counts_05.txt* that appears with the homework assignment. The file contains a list of 5-letter words (including names and proper nouns) and their counts from a large corpus of Wall Street Journal articles (roughly three million sentences). From the counts in this file compute the prior probability $P(w) = \text{COUNT}(w)/\text{COUNT}_{\text{total}}$. **As a sanity check, print out the ten <u>most</u> and <u>least</u> frequent 5-letter words. Do your results make sense?**

4

(b) Consider the following stages of the game. For each of the following, indicate the best next guess—namely, the letter ℓ that is most likely (probable) to be among the missing letters. Also report the probability $P(L_i = \ell$ for some $i \in \{1, 2, 3, 4, 5\}|E)$ for your guess ℓ. Your answers should fill in the last two columns of this table. (Some answers are shown so that you can check your work.)

| correctly guessed	incorrectly guessed	best next guess ℓ	$P(L_i = \ell$ for some $i \in \{1, 2, 3, 4, 5\}|E)$
− − − − −	{}		
− − − − −	{E, O}		
Q − − − −	{}		
Q − − − −	{U}		
− − Z E −	{A, D, I, R}		
− − − − −	{E, O}	I	0.6366
D − − I −	{}	A	0.8207
D − − I −	{A}	E	0.7521
− U − − −	{A, E, I, O, S}	Y	0.6270

(c) Include your source code when you submit to Gradescope. Do not forget the source code: it is worth many points on this assignment.

More fun: The hangman demo (linked in the schedule on the website) implements this program for words of length 6-10. You will also find count files for words of these lengths on in the folder of word counts, and if you modify your code to handle these different word lengths, you will also be able to check your answers against the demo. (This is totally optional, though.) Just to be perfectly clear, you are **not** required in this problem to implement a user interface or any general functionality for the game of hangman. You will only be graded on your word lists in (a), the completed table for (b), and your source code in (c).

2.4 Conditional independence

For the belief network shown below, indicate whether the following statements of marginal or conditional independence are **true (T)** or **false (F)**.

(a) _____ $\qquad P(F|H) \;=\; P(F|C,H)$

(b) _____ $\qquad P(E|A,B) \;=\; P(E|A,B,F)$

(c) _____ $\qquad P(E,F|B,G) \;=\; P(E|B,G)\,P(F|B,G)$

(d) _____ $\quad P(F|B,C,G,H) \;=\; P(F|B,C,E,G,H)$

(e) _____ $\quad P(A,B|D,E,F) \;=\; P(A,B|D,E,F,G,H)$

(f) _____ $\qquad P(D,E,F) \;=\; P(D)\,P(E|D)\,P(F|E)$

(g) _____ $\qquad P(A|F) \;=\; P(A)$

(h) _____ $\qquad P(E,F) \;=\; P(E)\,P(F)$

(i) _____ $\qquad P(D|A) \;=\; P(D|A,E)$

(j) _____ $\qquad P(B,C) \;=\; P(B)\,P(C)$
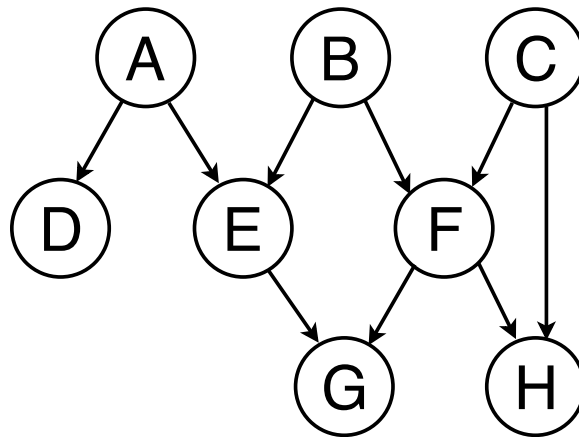
2.5 Subsets

For the DAG shown below, consider the following statements of marginal or conditional independence. Indicate the largest subset of nodes $\mathcal{S} \subseteq \{A, B, C, D, E, F\}$ for which each statement is true. Note that one possible answer is the empty set $\mathcal{S} = \emptyset$ or $\mathcal{S} = \{\}$ (whichever notation you prefer).

(a) $\qquad P(A) \quad = \quad P(A|\mathcal{S}) \qquad$ _____

(b) $\qquad P(A|C) \quad = \quad P(A|\mathcal{S}) \qquad$ _____

(c) $\qquad P(A|B,C) \quad = \quad P(A|\mathcal{S}) \qquad$ _____

(d) $\qquad P(B) \quad = \quad P(B|\mathcal{S}) \qquad$ _____

(e) $\qquad P(B|A,E) \quad = \quad P(B|\mathcal{S}) \qquad$ _____

(f) $\qquad P(B|A,C,E) \quad = \quad P(B|\mathcal{S}) \qquad$ _____

(g) $\qquad P(D) \quad = \quad P(D|\mathcal{S}) \qquad$ _____

(h) $\qquad P(D|A) \quad = \quad P(D|\mathcal{S}) \qquad$ _____

(i) $\qquad P(D|C,E) \quad = \quad P(D|\mathcal{S}) \qquad$ _____

(j) $\qquad P(D|F) \quad = \quad P(D|\mathcal{S}) \qquad$ _____