CSE 150A 250A. Homework 2

Due: *Mon Oct 13* (by 11:59 PM, Pacific Time, via gradescope)
Grace period: 24 hours

2.0 Basics

You should submit your homework assignments via gradescope:

`https://www.gradescope.com/courses/1132306`

You have two deliverables for this homework. Question 2.1-2.4 will be similar to homework 1 in which you must type your solutions and submit a pdf file to gradescope. For 2.5, fill in the starter code given and submit just your python file to the gradescope assignment: HW 2 - Coding Problem.

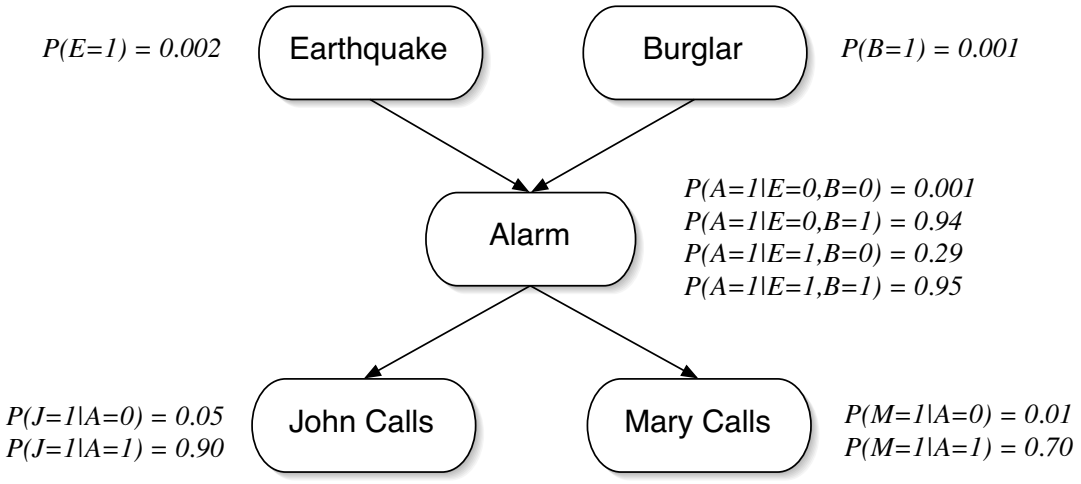
Late assignments will be accepted without penalty during the 24-hour grace period after the due date; however, such assignments may not be graded in a timely fashion. No assignments will be accepted beyond the grace period. Here is a primer on submitting PDF homework via gradescope:

`https://tinyurl.com/gradescope-guide`

If you have not done this before, please allow some extra time to familiarize yourself with this process.

2.1 Probabilistic inference (5 pts)

Recall the alarm belief network described in class. The directed acyclic graph (DAG) and conditional probability tables (CPTs) are shown below:



Compute numeric values for the following probabilities, exploiting relations of marginal and conditional independence as much as possible to simplify your calculations. You may re-use numerical results from lecture, but otherwise *show your work*. Be careful not to drop significant digits in your answer.

(a) $P(E=1|A=1)$
(b) $P(E=1|A=1,B=0)$
(c) $P(A=1|M=1)$
(d) $P(A=1|M=1,J=0)$
(e) $P(A=1|M=0)$
(f) $P(A=1|M=0,B=1)$

Consider your results in (b) versus (a), (d) versus (c), and (f) versus (e). Do they seem consistent with commonsense patterns of reasoning?

Solution 2.1

Known data

$P(E=1)=0.002, \quad P(E=0)=0.998, \quad P(B=1)=0.001, \quad P(B=0)=0.999,$
$P(A=1 \mid E=1,B=1)=0.95,\ P(A=1 \mid E=1,B=0)=0.29,$
$P(A=1 \mid E=0,B=1)=0.94,\ P(A=1 \mid E=0,B=0)=0.001,$
$P(J=1 \mid A=1)=0.9,\ P(J=1 \mid A=0)=0.05,\ \Rightarrow\ P(J=0 \mid A=1)=0.1,\ P(J=0 \mid A=0)=0.95,$
$P(M=1 \mid A=1)=0.7,\ P(M=1 \mid A=0)=0.01,\ \Rightarrow\ P(M=0 \mid A=1)=0.3,\ P(M=0 \mid A=0)=0.99.$

(a) $P(E=1 \mid A=1)$

We need: $P(A=1 \mid E), P(E)$ and $P(A=1 \mid E,B), P(B)$.

$$P(E=1 \mid A=1)=\frac{P(A=1 \mid E=1)\,P(E=1)}{P(A=1)}$$

$$\begin{aligned}P(A=1 \mid E=1) &= P(A=1 \mid E=1,B=1)P(B=1)+P(A=1 \mid E=1,B=0)P(B=0)\\ &= 0.95\cdot 0.001+0.29\cdot 0.999=0.29066\end{aligned}$$

$$\begin{aligned}P(A=1 \mid E=0) &= P(A=1 \mid E=0,B=1)P(B=1)+P(A=1 \mid E=0,B=0)P(B=0)\\ &= 0.94\cdot 0.001+0.001\cdot 0.999=0.00194\end{aligned}$$

$$\begin{aligned}P(A=1) &= P(A=1 \mid E=1)P(E=1)+P(A=1 \mid E=0)P(E=0)\\ &= 0.29066\cdot 0.002+0.00194\cdot 0.998=0.00252\end{aligned}$$

Substitute:

$$P(E=1 \mid A=1)=\frac{0.29066\cdot 0.002}{0.00252}=0.23101$$

(b) $P(E=1 \mid A=1,B=0)$

We need: $P(E), P(A=1 \mid E,B=0)$.

$$P(E=1 \mid A=1,B=0)=\frac{P(A=1 \mid E=1,B=0)P(E=1)}{P(A=1 \mid E=1,B=0)P(E=1)+P(A=1 \mid E=0,B=0)P(E=0)}$$

Substitute:

$$P(E=1 \mid A=1,B=0)=\frac{0.29\cdot 0.002}{0.29\cdot 0.002+0.001\cdot 0.998}=\frac{0.00058}{0.001578}=0.36755$$

(c) $P(A = 1 \mid M = 1)$

We need: $P(M = 1 \mid A)$ and $P(A)$.

$$P(A = 1 \mid M = 1) = \frac{P(M = 1 \mid A = 1)P(A = 1)}{P(M = 1)}$$

$$\begin{aligned} P(M = 1) &= P(M = 1 \mid A = 1)P(A = 1) + P(M = 1 \mid A = 0)P(A = 0) \\ &= 0.7 \cdot 0.00252 + 0.01 \cdot 0.99748 = 0.01174 \end{aligned}$$

Substitute:

$$P(A = 1 \mid M = 1) = \frac{0.7 \cdot 0.00252}{0.01174} = 0.15009$$

(d) $P(A = 1 \mid M = 1, J = 0)$

We need: $P(M \mid A), P(J \mid A), P(A)$.

$$P(A = 1 \mid M = 1, J = 0) = \frac{P(M = 1 \mid A = 1)\, P(J = 0 \mid A = 1)\, P(A = 1)}{\sum_{a \in \{0,1\}} P(M = 1 \mid A = a)\, P(J = 0 \mid A = a)\, P(A = a)}$$

$$\begin{aligned} P(M = 1, J = 0 \mid A = 1) &= 0.7 \cdot 0.1 = 0.07 \\ P(M = 1, J = 0 \mid A = 0) &= 0.01 \cdot 0.95 = 0.0095 \\ P(M = 1, J = 0) &= 0.07 \cdot 0.00252 + 0.0095 \cdot 0.99748 = 0.00965 \end{aligned}$$

Substitute:

$$P(A = 1 \mid M = 1, J = 0) = \frac{0.07 \cdot 0.00252}{0.00965} = 0.01825$$

(e) $P(A = 1 \mid M = 0)$

We need: $P(M \mid A), P(A)$.

$$P(A = 1 \mid M = 0) = \frac{P(M = 0 \mid A = 1)P(A = 1)}{P(M = 0)}$$

$$\begin{aligned} P(M = 0) &= P(M = 0 \mid A = 1)P(A = 1) + P(M = 0 \mid A = 0)P(A = 0) \\ &= 0.3 \cdot 0.00252 + 0.99 \cdot 0.99748 = 0.98826 \end{aligned}$$

Substitute:

$$P(A = 1 \mid M = 0) = \frac{0.3 \cdot 0.00252}{0.98826} = 0.00076$$

(f) $P(A = 1 \mid M = 0, B = 1)$

We need: $P(A \mid B = 1)$ and $P(M \mid A)$.

$$\begin{aligned} P(A = 1 \mid B = 1) &= P(A = 1 \mid E = 1, B = 1)P(E = 1) + P(A = 1 \mid E = 0, B = 1)P(E = 0) \\ &= 0.95 \cdot 0.002 + 0.94 \cdot 0.998 = 0.94002 \end{aligned}$$

$$P(A = 0 \mid B = 1) = 1 - 0.94002 = 0.05998$$

$$P(A = 1 \mid M = 0, B = 1) = \frac{P(M = 0 \mid A = 1)P(A = 1 \mid B = 1)}{P(M = 0 \mid A = 1)P(A = 1 \mid B = 1) + P(M = 0 \mid A = 0)P(A = 0 \mid B = 1)}$$

Substitute:

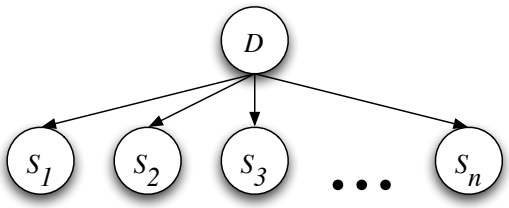
$$P(A = 1 \mid M = 0, B = 1) = \frac{0.3 \cdot 0.94002}{0.34139} = 0.82606$$

Comparison and commonsense check

- (b) vs (a): $0.36755 > 0.23101$. When burglary is ruled out, an alarm makes earthquake more likely, because earthquake is the only remaining plausible parent. This matches our everyday reasoning about causal explanations.
- (d) vs (c): $0.01825 \ll 0.15009$. Mary calling but John not calling is inconsistent evidence for a true alarm, so the posterior drops sharply. This reflects the intuition that conflicting reports undermine our belief. The decrease is large, which is consistent with John being a reliable witness in the model.
- (f) vs (e): $0.82606 \gg 0.00076$. With no burglary information, an alarm is almost impossible if Mary does not call. But when burglary is known to be true, the alarm becomes very likely even if Mary is silent. This shows how strong causal evidence (burglary) dominates weaker evidence (Mary's phone call).

2.2 Probabilistic reasoning (5 pts)

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 $(D{=}0)$, the first symptom occurs with probability one,

$$P(S_1{=}1|D{=}0) = 1,$$

while the k^{th} symptom (with $k \geq 2$) occurs with probability

$$P(S_k{=}1|D{=}0) = \frac{f(k-1)}{f(k)},$$

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

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

By contrast, in the other form of the disease $(D{=}1)$, all the symptoms are uniformly likely to be observed, with

$$P(S_k{=}1|D{=}1) = \frac{1}{2}$$

for all k. Suppose that on the k^{th} day of the month, a test is done to determine whether the patient is exhibiting the k^{th} symptom, and that each such test returns a positive result. Thus, on the k^{th} 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.

Solution 2.2

(a) We need: the posterior odds $r_k = \dfrac{P(D=0 \mid S_1{=}1, \ldots, S_k{=}1)}{P(D=1 \mid S_1{=}1, \ldots, S_k{=}1)}$.

$$\begin{aligned} r_k &= \frac{P(D=0)\prod_{i=1}^{k} P(S_i=1 \mid D=0)}{P(D=1)\prod_{i=1}^{k} P(S_i=1 \mid D=1)} \\ &= \frac{P(D=0)}{P(D=1)} \cdot \frac{P(S_1=1 \mid D=0)\prod_{i=2}^{k} \frac{f(i-1)}{f(i)}}{\prod_{i=1}^{k} \frac{1}{2}} \\ &= \frac{P(D=0)}{P(D=1)} \cdot \frac{1 \cdot \left(\frac{f(1)}{f(2)} \cdot \frac{f(2)}{f(3)} \cdots \frac{f(k-1)}{f(k)}\right)}{(\frac{1}{2})^k} \\ &= \frac{P(D=0)}{P(D=1)} \cdot \frac{f(1)}{f(k)} \cdot 2^k. \end{aligned}$$

Substitute $P(D=0) = P(D=1) = \frac{1}{2}$ and $f(1) = 2^1 + (-1)^1 = 1$:

$$r_k = \frac{2^k}{f(k)} = \frac{2^k}{2^k + (-1)^k}.$$

Diagnosis:

$$\text{if } k \text{ is odd, } r_k = \frac{2^k}{2^k - 1} > 1 \Rightarrow D = 0; \qquad \text{if } k \text{ is even, } r_k = \frac{2^k}{2^k + 1} < 1 \Rightarrow D = 1.$$

Which means: The doctor makes his diagnosis depend on whether the day of the month is odd or even. If the day is odd, then the disease should be in the form of $D = 0$, otherwise, it should be in the form of $D = 1$.

(b) The diagnosis becomes less certain as more symptoms are observed.

According to the deviation of r_k from 1 as k grows.

$$\begin{aligned} r_k &= \frac{2^k}{2^k + (-1)^k} = 1 - \frac{(-1)^k}{2^k + (-1)^k} \\ \Rightarrow \quad |r_k - 1| &= \frac{1}{2^k + (-1)^k}. \end{aligned}$$

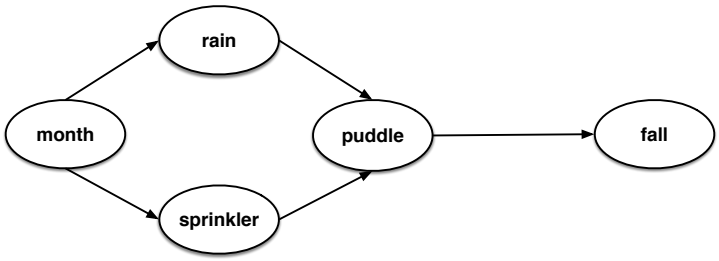
We can have:

$$lim_{k\to\infty}|r_k - 1| = 0.$$

As $k \to \infty$, $|r_k - 1| \to 0$ and hence $r_k \to 1$ (odd k from above, even k from below), which means the difference between r_k and 1 is becoming less and less as k increases. Therefore the diagnosis becomes less certain with more symptoms, even though the favored class flips with parity.

2.3 Conditional independence (5 pts)

Consider the DAG shown below, describing the following domain. Given the `month` of the year, there is some probability of `rain`, and also some probability that the `sprinkler` is turned on. Either of these events leads to some probability that a `puddle` forms on the sidewalk, which in turn leads to some probability that someone has a `fall`.



List all the conditional independence relations that must hold in any probability distribution represented by this DAG. More specifically, list all tuples $\{X, Y, E\}$ such that $P(X, Y|E) = P(X|E)P(Y|E)$, where

$$\begin{aligned} X, Y &\in \{\texttt{month}, \texttt{rain}, \texttt{sprinkler}, \texttt{puddle}, \texttt{fall}\}, \\ E &\subseteq \{\texttt{month}, \texttt{rain}, \texttt{sprinkler}, \texttt{puddle}, \texttt{fall}\}, \\ X &\neq Y, \\ X, Y &\notin E. \end{aligned}$$

Hint: There are sixteen such tuples, not counting those that are equivalent up to exchange of X and Y. Do any of the tuples contain the case $E = \emptyset$?

Solution 2.3

rain $\perp$ sprinkler

(rain, sprinkler, {month})

puddle $\perp$ month

(puddle, month, {rain, sprinkler})
(puddle, month, {rain, sprinkler, fall})

fall $\perp$ month

(fall, month, {puddle})
(fall, month, {puddle, rain})
(fall, month, {puddle, sprinkler})
(fall, month, {puddle, rain, sprinkler})
(fall, month, {rain, sprinkler})

fall $\perp$ rain

(fall, rain, {puddle})
(fall, rain, {puddle, month})
(fall, rain, {puddle, sprinkler})
(fall, rain, {puddle, month, sprinkler})

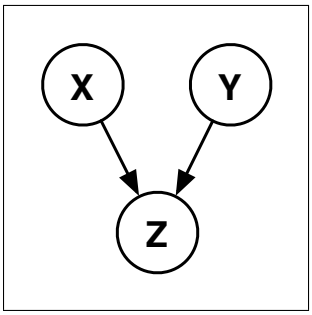
fall $\perp$ sprinkler

(fall, sprinkler, {puddle})
(fall, sprinkler, {puddle, month})
(fall, sprinkler, {puddle, rain})
(fall, sprinkler, {puddle, month, rain})

Answer of Question in Hint:

There are sixteen distinct conditional independence relations (not counting those that are equivalent up to exchange of X and Y). No relation holds with $E = \emptyset$.

2.4 Noisy-OR (5 pts)



Nodes:	$X \in \{0,1\}, Y \in \{0,1\}, Z \in \{0,1\}$
Noisy-OR CPT:	$P(Z=1\|X,Y) = 1-(1-p_x)^X(1-p_y)^Y$
Parameters:	$p_x \in [0,1], p_y \in [0,1], p_x < p_y$

Suppose that the nodes in this network represent binary random variables and that the CPT for $P(Z|X,Y)$ is parameterized by a noisy-OR model, as shown above. Suppose also that

$$0 < P(X=1) < 1,$$
$$0 < P(Y=1) < 1,$$

while the parameters of the noisy-OR model satisfy:

$$0 < p_x < p_y < 1.$$

Consider the following pairs of probabilities. In each case, indicate whether the probability on the left is equal $(=)$, greater than $(>)$, or less than $(<)$ the probability on the right. The first one has been filled in for you as an example. (You should use your intuition for these problems; you are **not** required to show work.)

	$P(X=1)$	$\boxed{=}$	$P(X=1)$
(a)	$P(Z=1 \mid X=0,Y=0)$	$<$	$P(Z=1 \mid X=0,Y=1)$
(b)	$P(Z=1 \mid X=1,Y=0)$	$<$	$P(Z=1 \mid X=0,Y=1)$
(c)	$P(Z=1 \mid X=1,Y=0)$	$<$	$P(Z=1 \mid X=1,Y=1)$
(d)	$P(X=1)$	$<$	$P(X=1 \mid Z=1)$
(e)	$P(X=1)$	$=$	$P(X=1 \mid Y=1)$
(f)	$P(X=1 \mid Z=1)$	$>$	$P(X=1 \mid Y=1,Z=1)$
(g)	$P(X=1)\,P(Y=1)\,P(Z=1)$	$<$	$P(X=1,Y=1,Z=1)$

2.5 Programming Question: Hangman (15 pts)

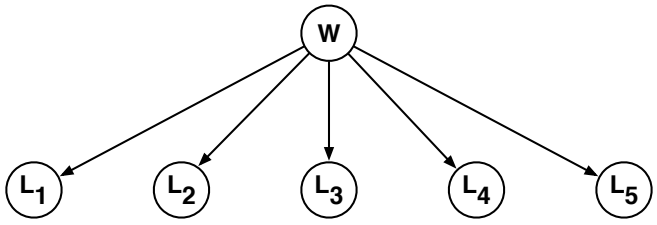
Consider the belief network shown below, where the random variable W stores a five-letter word and the random variable $L_i \in \{\texttt{A}, \texttt{B}, \ldots, \texttt{Z}\}$ reveals only the word's ith letter. Also, suppose that these five-letter words are chosen at random from a large corpus of text 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 (A through 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 D, I, M—you've learned that the letter I does *not* appear, and that the letters D and M appear as follows:

M ___ D ___ 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 works, 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:

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

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:

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

And the first term we obtain from Bayes rule:

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

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

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:

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

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 implement both of these calculations. **This assignment is to be done in Python and submitted gradescope separately.**

This assignment will be autograded. Your code will be evaluated based off the the run() function, which should return 3 deliverables: a set containing the 15 most frequent words, a set containing the 14 least frequent words, and a dataframe containing the best next guess and associated probability for a given state. When you submit, the autograder will only show whether your code compiled and if it passed the visible test cases for part (b). Thus, it is up to you to make sure your code works for the hidden test cases.

(a) Download the file *hw1_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 $P(w) = \text{COUNT}(w) / \sum_{w'} \text{COUNT}(w')$. **As a sanity check, print out the fifteen most frequent 5-letter words, as well as the fourteen least frequent 5-letter words. Do your results make sense?**

(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 $P(L_i = \ell \text{ for some } i \in \{1, 2, 3, 4, 5\} | E)$ 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 $i \in \{1, 2, 3, 4, 5\} \mid E)$
_ _ _ _ _	{}		
_ _ _ _ _	{E, A}		
A _ _ _ S	{}		
A _ _ _ S	{I}	E	0.7127
_ _ O _ _	{A, E, M, N, T}		
_ _ _ _ _	{E, O}	I	0.6366
D _ _ I _	{}		
D _ _ I _	{A}	E	0.7521
_ U _ _ _	{A, E, I, O, S}		