

---

## CSE 250A. Assignment 2

**Out:** Tue Oct 10

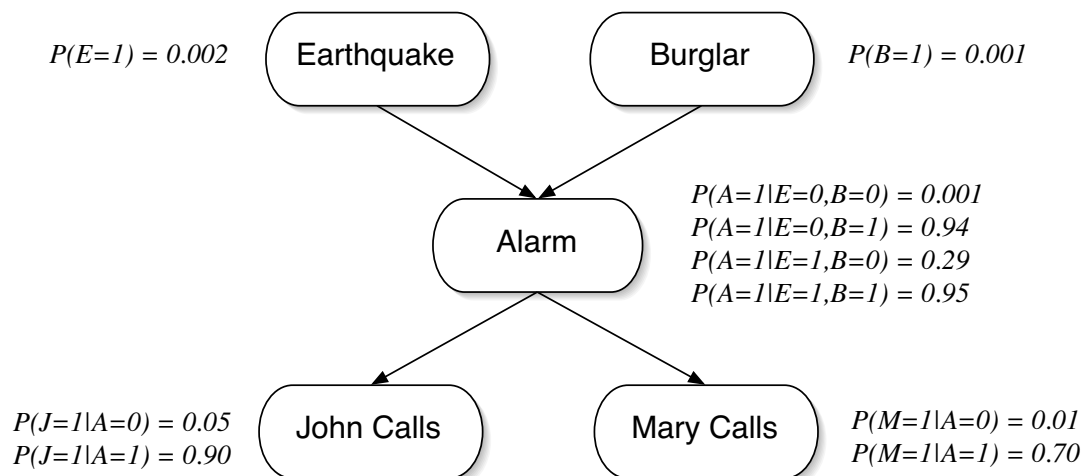
**Due:** Tue Oct 17 (in class)

**Reading:** Russell & Norvig, Chapters 13-14.

---

### 2.1 Probabilistic inference

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

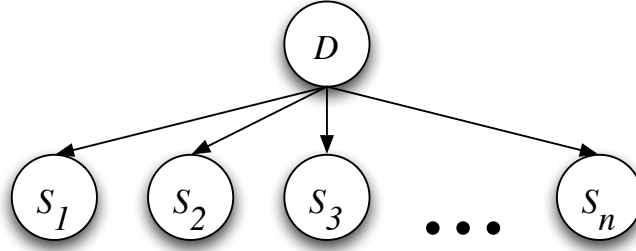
Consider your results in (b) versus (a), (d) versus (c), and (f) versus (e). Do they seem consistent with commonsense patterns of reasoning?

---

---

## 2.2 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 the first form of the disease ( $D=0$ ), the first symptom occurs with probability one,

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

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

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

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.

- 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.
  - Does the diagnosis become more or less certain as more symptoms are observed? Explain.
-

---

## 2.3 Sigmoid function

Let  $Y \in \{0, 1\}$  denote a binary random variable that depends on  $k$  other random variables  $X_i$  as:

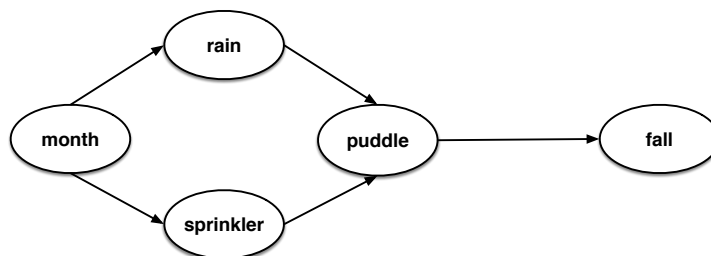
$$P(Y=1|X_1=x_1, X_2=x_2, \dots, X_k=x_k) = \sigma\left(\sum_{i=1}^k w_i x_i\right) \quad \text{where} \quad \sigma(z) = \frac{1}{1 + e^{-z}}.$$

The real-valued parameters  $w_i$  in this CPT are known as weights. The so-called sigmoid function  $\sigma(z)$  arises in many contexts. In neural networks, it models the probability that a neuron  $Y$  fires given its input from other neurons  $X_i$ ; the weights  $w_i$  describe the connections between neurons. In statistics, the sigmoid function appears in models of logistic regression. Sketch the function  $\sigma(z)$ , and verify the following properties:

- (a)  $\sigma'(z) = \sigma(z)\sigma(-z)$ .
  - (b)  $\sigma(-z) + \sigma(z) = 1$ .
  - (c)  $L(\sigma(z)) = z$ , where  $L(p) = \log\left(\frac{p}{1-p}\right)$  is the *log-odds* function.
  - (d)  $w_i = L(p_i)$ , where  $p_i = P(Y=1|X_i=1, X_j=0 \text{ for all } j \neq i)$ .
- 

## 2.4 Conditional independence

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

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

**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  $E = \emptyset$ ?

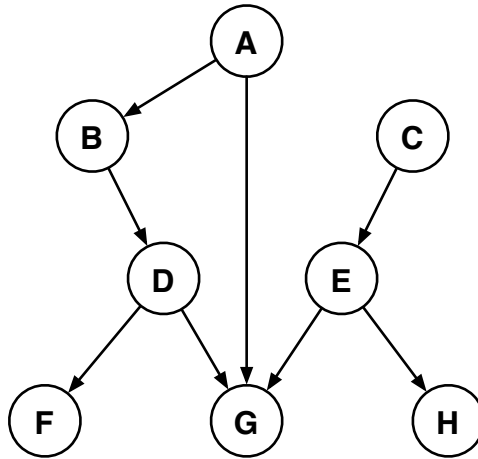
---

The Markov blanket  $B_X$  of a node  $X$  in a belief network includes its parents, its children, and the parents of its children (excluding the node  $X$  itself). Draw a picture of a Markov blanket; then, by appealing to the three conditions for d-separation, prove that for any node  $Y$  outside the Markov blanket of  $X$  (that is, satisfying  $Y \notin B_X$  and  $Y \neq X$ ), we have:

---

## 2.6 True or false

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



\_\_\_\_\_

$$P(B|G, C) = P(B|G)$$

\_\_\_\_\_

$$P(F, G|D) = P(F|D) P(G|D)$$

\_\_\_\_\_

$$P(A, C) = P(A) P(C)$$

\_\_\_\_\_

$$P(D|B, F, G) = P(D|B, F, G, A)$$

\_\_\_\_\_

$$P(F, H) = P(F) P(H)$$

\_\_\_\_\_

$$P(D, E|F, H) = P(D|F) P(E|H)$$

\_\_\_\_\_

$$P(F, C|G) = P(F|G) P(C|G)$$

\_\_\_\_\_

$$P(D, E, G) = P(D) P(E) P(G|D, E)$$

\_\_\_\_\_

$$P(H|C) = P(H|A, B, C, D, F)$$

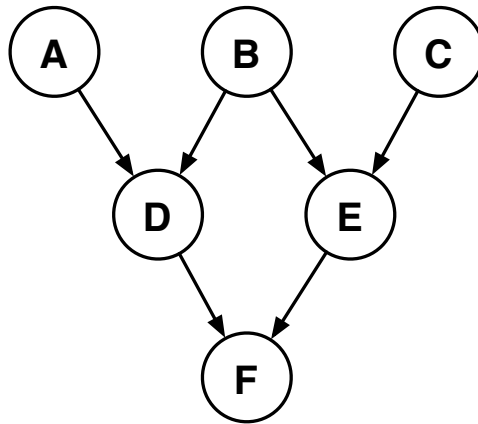
\_\_\_\_\_

$$P(H|A, C) = P(H|A, C, G)$$

---

---

## 2.7 Subsets



For the belief network shown above, consider the following statements of conditional independence. Indicate the largest subset of nodes  $\mathcal{S} \subset \{A, B, C, D, E, F\}$  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). The first one is done as an example.

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

$$P(A|D) = P(A|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

$$P(A|B, D) = P(A|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

$$P(B|D, E) = P(B|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

$$P(E) = P(E|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

$$P(E|F) = P(E|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

$$P(E|D, F) = P(E|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

$$P(E|B, C) = P(E|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

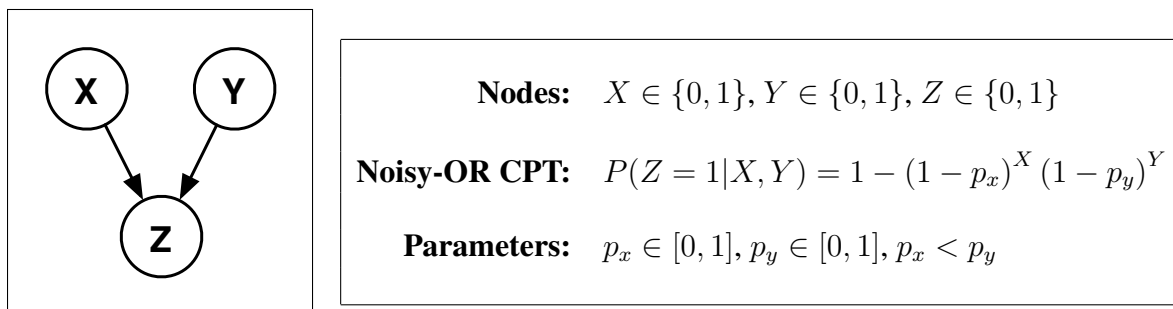
$$P(F) = P(F|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

$$P(F|D) = P(F|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

$$P(F|D, E) = P(F|\mathcal{S}) \quad \underline{\hspace{2cm}}$$

---

## 2.8 Noisy-OR



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:

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

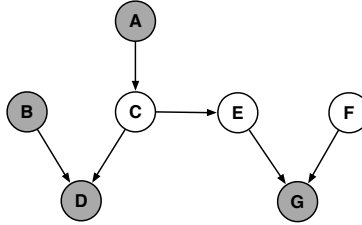
Consider the following pairs of probabilities. In each case, indicate whether the probability on the left is equal ( $=$ ), greater than ( $>$ ), or less than ( $<$ ) the probability on the right. The first one has been filled in for you as an example. (You should use your intuition for these problems;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

	$P(X=1)$	<input checked="" style="width: 40px; height: 25px; border: 1px solid black;" type="text" value="="/>	$P(X=1)$
(a)	$P(Z=1 X=0, Y=0)$	<input style="width: 40px; height: 25px; border: 1px solid black;" type="text"/>	$P(Z=1 X=0, Y=1)$
(b)	$P(Z=1 X=1, Y=0)$	<input style="width: 40px; height: 25px; border: 1px solid black;" type="text"/>	$P(Z=1 X=0, Y=1)$
(c)	$P(Z=1 X=1, Y=0)$	<input style="width: 40px; height: 25px; border: 1px solid black;" type="text"/>	$P(Z=1 X=1, Y=1)$
(d)	$P(X=1)$	<input style="width: 40px; height: 25px; border: 1px solid black;" type="text"/>	$P(X=1 Z=1)$
(e)	$P(X=1)$	<input style="width: 40px; height: 25px; border: 1px solid black;" type="text"/>	$P(X=1 Y=1)$
(f)	$P(X=1 Z=1)$	<input style="width: 40px; height: 25px; border: 1px solid black;" type="text"/>	$P(X=1 Y=1, Z=1)$
(g)	$P(X=1) P(Y=1) P(Z=1)$	<input style="width: 40px; height: 25px; border: 1px solid black;" type="text"/>	$P(X=1, Y=1, Z=1)$

*Challenge (optional):* for each case, prove rigorously the correctness of your answer.

---

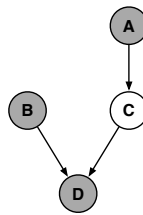
## 2.9 Polytree inference



Consider the belief network shown above. In this problem you will be guided through an efficient computation of the posterior probability  $P(F|A, B, D, G)$ . You should perform these computations *efficiently*—that is, by exploiting the structure of the DAG and not marginalizing over more variables than necessary.

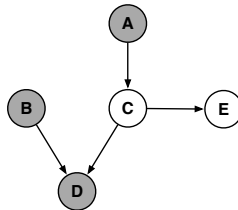
(a) **Bayes rule**

Show how to compute the posterior probability  $P(C|A, B, D)$  in terms of the conditional probability tables (CPTs) for these nodes—i.e., in terms of  $P(A)$ ,  $P(B)$ ,  $P(C|A)$ , and  $P(D|B, C)$ . For this problem, it may help to consider just the part of the network shown below.



(b) **Marginalization**

Show how to compute the posterior probability  $P(E|A, B, D)$  in terms of your answer from part (a) and the CPTs of the belief network. For this problem, it may help to consider just the part of the network shown below.



(c) **Marginalization**

Show how to compute the posterior probability  $P(G|A, B, D)$  in terms of your answer from part (b) and the CPTs of the belief network.

(d) **Explaining away**

Show how to compute the posterior probability  $P(F|A, B, D, G)$  in terms of your answers from parts (b,c) and the CPTs of the belief network.

---