

# CS760 Spring 2019 Homework 2

Due Mar 7 at 11:59pm

Name: Stewart Kerr

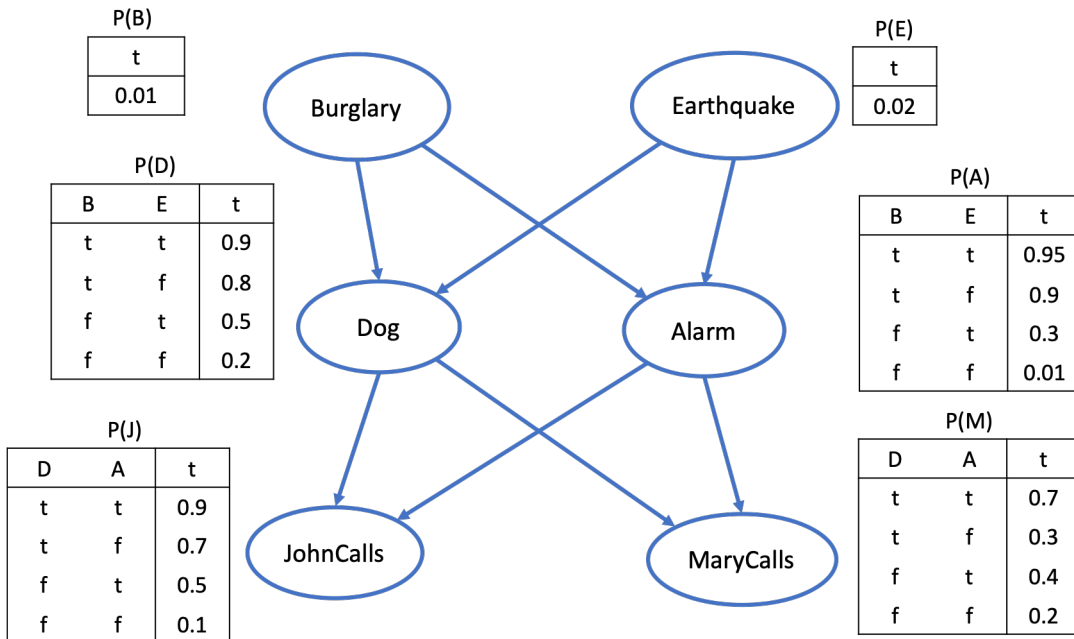
Email: shkerr@wisc.edu

## Written Problems

**NOTE:** For the following written problems, put your answer in `hw2.pdf`. You are required to provide detailed solutions including the intermediate results for each step. Otherwise, you will not get full credit. You can also add figures or tables whenever necessary. If your solutions are handwritten, make sure they are legible.

- (8 pts) Suppose you have a Bayesian network with 6 binary random variables shown as follows, where  $t$  and  $f$  stand for *true* and *false* respectively.

Compute the probability:  $P(d|b, \neg a, j, m)$ .



From the formula of conditional probability, we know that  $P(d|b, \neg a, j, m) = \frac{P(d, b, \neg a, j, m)}{P(b, \neg a, j, m)}$ .

By the law of total probability, we can account for both earthquake outcomes and rewrite this fraction as

$$P(d|b, \neg a, j, m) = \frac{P(e, d, b, \neg a, j, m) + P(\neg e, d, b, \neg a, j, m)}{P(e, d, b, \neg a, j, m) + P(\neg e, d, b, \neg a, j, m) + P(e, \neg d, b, \neg a, j, m) + P(\neg e, \neg d, b, \neg a, j, m)}$$

Then, to get each probability in this sum, we just multiply the corresponding entries of the CPT (taking outcomes of dependent variables into account). Thus,

$$P(e, d, b, \neg a, j, m) = P(e) \times P(b) \times P(d|e, b) \times P(\neg a|e, b) \times P(j|d, \neg a) \times P(m, d, \neg a)$$

$$= 0.02 \times 0.01 \times 0.9 \times 0.05 \times 0.7 \times 0.3 = 1.89 \times 10^{-6}$$

$$P(\neg e, d, b, \neg a, j, m) = 0.98 \times 0.01 \times 0.8 \times 0.1 \times 0.7 \times 0.3 = 0.00165$$

$$P(e, \neg d, b, \neg a, j, m) = 0.02 \times 0.01 \times 0.1 \times 0.05 \times 0.1 \times 0.2 = 2 \times 10^{-8}$$

$$P(\neg e, \neg d, b, \neg a, j, m) = 0.98 \times 0.01 \times 0.2 \times 0.1 \times 0.1 \times 0.2 = 3.92 \times 10^{-6}$$

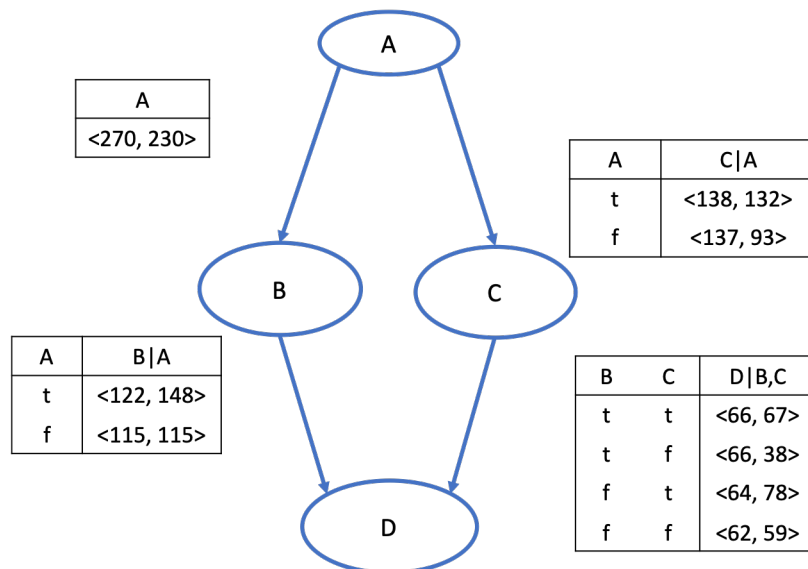
Now, we plug all of those numbers into our equation for  $P(d|b, \neg a, j, m)$ .

$$P(d|b, \neg a, j, m) = \frac{1.89 \times 10^{-6} + 0.00165}{1.89 \times 10^{-6} + 0.00165 + 2 \times 10^{-8} + 3.92 \times 10^{-6}} = 0.977$$

Intuitively, this makes sense. If both John and Mary call, but the alarm did *not* sound, then that places a higher "burden" on the dog barking to be the reason behind their calling. Thus, the probability that the dog barked will be higher.

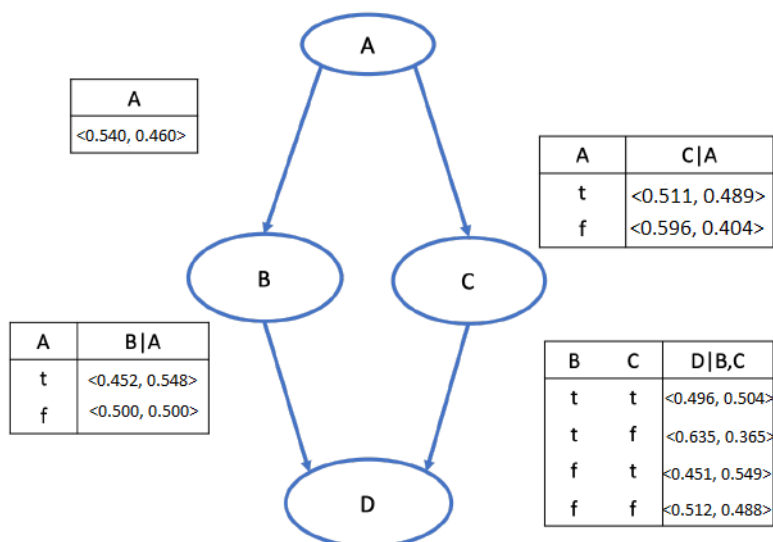
2. Given the following Bayesian network and sample counts in each table, where sample counts  $\langle n_{\text{true}}, n_{\text{false}} \rangle$ , there are  $n_{\text{true}}$  samples with true labels and  $n_{\text{false}}$  samples with false labels for this attribute. For example,  $\langle 138, 132 \rangle$  in table  $C|A$  says given the condition of  $A = \text{true}$ , there are 138 instances are true and 132 are false with regard to attribute  $C$ .

You need to answer the following two questions.



- (a) (2 pts) Construct the conditional probability tables (CPTs) based on the above sample count tables, using maximum likelihood estimation. You need to both show the true probability  $P_{\text{true}}$  and false probability  $P_{\text{false}}$  for each case, and organize them in the format of  $\langle P_{\text{true}}, P_{\text{false}} \rangle$ . For example, for the case  $Y|X_1, X_2$ , your answer will look like  $\langle P(Y|X_1, X_2), P(\neg Y|X_1, X_2) \rangle$ . Keep **at least 3 digits of precision**. (You may reuse the same structure as the above tables, just plugging in the conditional probabilities in the place of sample counts. For more information, please refer to the lecture notes `BNS-1.pdf`)

See the figure below for the CPT tables.



- (b) (10 pts) Show the result of one cycle of the EM algorithm to update the CPTs you derived in step (a), using 10 another instances with  $A=\text{true}$ ,  $B=\text{false}$ ,  $C=?$ , and  $D=\text{true}$  ('?' means missing value). Keep **at least 2 digits of precision**.

The EM algorithm has two steps -

E) Using the current CPT, calculate expected values of the missing data.

M) Update the model using the imputed values from E which maximize probability of the data

**Expectation step** Given the CPT tables constructed above, we have

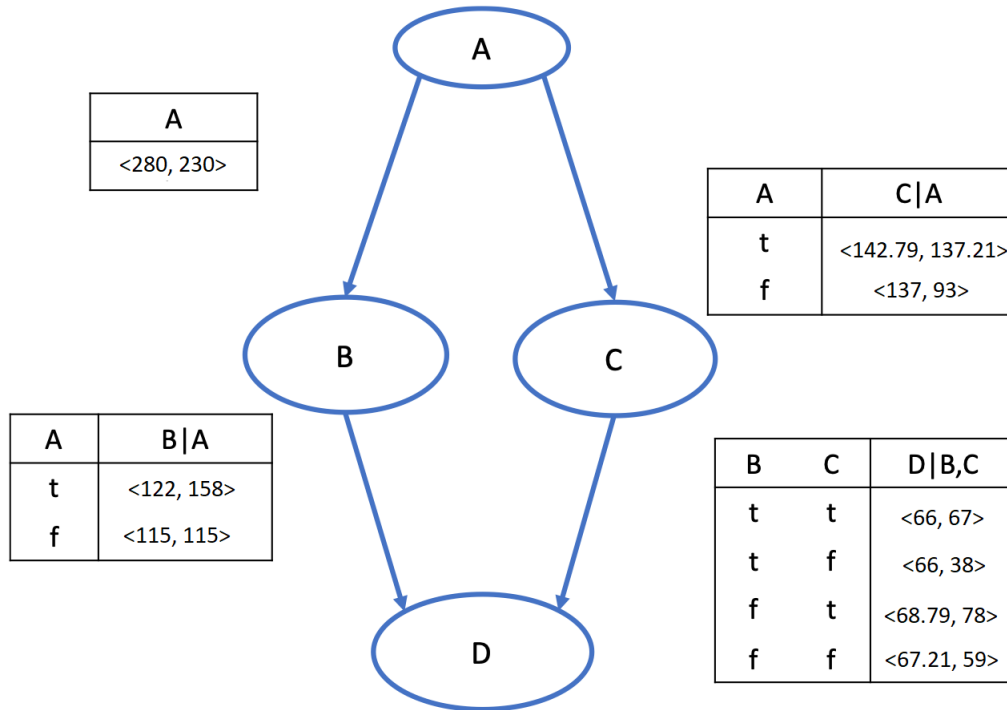
$$P(c|a, \neg b, d) = \frac{P(a, \neg b, c, d)}{P(a, \neg b, c, d) + P(a, \neg b, \neg c, d)} = \frac{0.540 \times 0.548 \times 0.511 \times 0.451}{(0.540 \times 0.548 \times 0.511 \times 0.451) + (0.540 \times 0.548 \times 0.489 \times 0.512)} = 0.479$$

and

$$P(\neg c|a, \neg b, d) = 1 - P(c|a, \neg b, d) = 1 - 0.479 = 0.521$$

Thus, the expected counts for  $c$  and  $\neg c$  using another 10 instances are  $10 \times 0.479 = 4.79$  and  $10 \times 0.521 = 5.21$  respectively.

**Maximization step** In the maximization step, we add the new data (with expected counts) to the previous counts and then recalculate the CPT table. After adding the new data, the counts table is:



Thus, the new CPT table is:

