

**Guidelines.** Students may discuss problems in the assignment across group boundaries. However, each group must write down the answers independently. Submit as a SINGLE PDF (either typeset or scanned) per group. IMPORTANT: name your PDF file as

NETID.PDF

where NETID is your actual netid. For groups of two students, the file name should be

NETID1-NETID2.PDF

One group should submit only a single copy. Submitting two copies through two students accounts for the same group will result in 20% penalty for both students.

**Problem 1 [10 points].** Marie is getting married tomorrow at an outdoor ceremony in NYC. In recent years, it rained about once every three days. The weatherman has predicted there will be no rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 80% of the time. When it doesn't rain, he incorrectly forecasts rain 20% of the time. What is the probability that it will rain on the day of Marie's wedding according to Bayes' rule?

**Problem 2 [20 points].** Consider the Monty Hall problem that we discussed in class. Instead of having three doors with one door having a price, now suppose there are four doors with two doors hiding goats behind them and one door having a price behind it. The other door has nothing behind it. To play the game, the guest will pick a door and the host will then reveal another door with a goat behind. The guest can then choose whether to stick to the original choice or switching. Should the guest switch? What is the probability of winning if the guest switches?

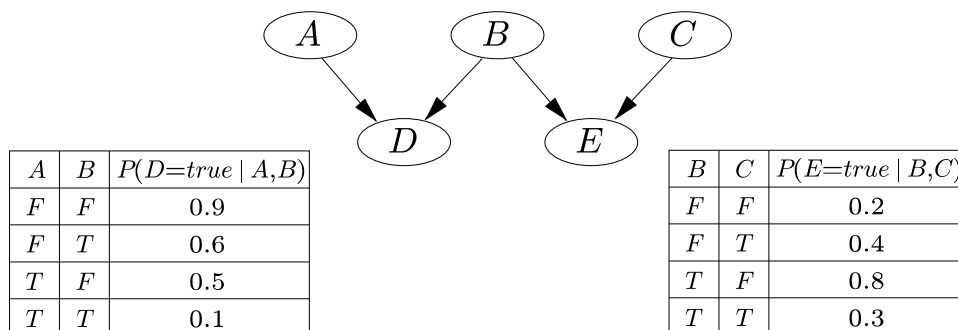
**Problem 3 [20 points].** Using the sampling procedure covered in class and python (or any other way you see fit), generate  $N$  samples for the cumulative distribution function

$$\Phi(x) = \frac{1}{2} + \frac{\text{sign}(x)}{2} \sqrt{1 - e^{-\frac{2x^2}{\pi}}}.$$

$\text{sign}(x)$  is the sign of  $x$ , e.g.,  $\text{sign}(-1.6) = -1$ ,  $\text{sign}(5.3) = 1$ , and  $\text{sign}(0) = 0$ . Discard any sample if  $|x| > 5$ . Plot a histogram of your data from  $-5$  to  $5$  with  $0.2$  increments (i.e., you should have 50 bins). Do this for  $N = 100, 200, 500$ , and  $1000$ . You should submit four figures. Note that you can easily do histograms in python using matplotlib. (Hint: it can be slightly tricky to compute  $x$  from the CDF; but you don't need to).

**Problem 4 [30 points].** Consider the following Bayesian network, where variables  $A$  through  $E$  are all Boolean valued:

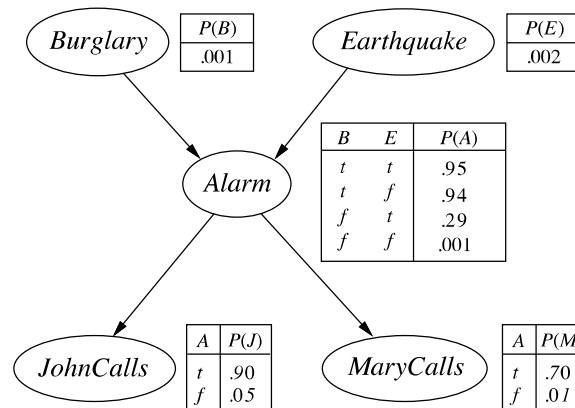
$$P(A=\text{true})=0.2 \quad P(B=\text{true})=0.5 \quad P(C=\text{true})=0.8$$



a) [5 points] What is the probability that all five of these Boolean variables are true?

- b) [10 points] Compute  $P(A, E \mid B)$ .
- c) [5 points] Compute  $P(A = \text{true} \mid B = \text{false}, C = \text{false}, D = \text{false}, E = \text{false})$ .
- d) [5 points] Compute  $P(C = \text{true}, D = \text{false} \mid B = \text{false})$ .
- e) [5 points] Compute  $P(B = \text{true}, C = \text{false}, D = \text{true} \mid A = \text{false})$ .

**Problem 5 [20 points].** For this problem, check the Variable Elimination algorithm in your book. Also consider the Bayesian network from the “burglary” example.



- a) [5 points] Apply variable elimination to the query:

$$P(\text{Burglary} \mid \text{JohnsCalls} = \text{true}, \text{MaryCalls} = \text{true})$$

and show in detail the calculations that take place. Use your book to confirm that your answer is correct.

- b) [5 points] Count the number of arithmetic operations performed (additions, multiplications, divisions), and compare it against the number of operations performed by the tree enumeration algorithm.
- c) [10 points] Suppose a Bayesian network has the form of a *chain*: a sequence of Boolean variables  $X_1, \dots, X_n$  where  $\text{Parents}(X_i) = \{X_{i-1}\}$  for  $i = 2, \dots, n$ . What is the complexity of computing  $P(X_1 \mid X_n = \text{true})$  using enumeration (in Big  $O$  notation)? What is the complexity with variable elimination?