COMPSCI 402 Assignment 4

Q1. Probability 3-point

(a) A, B, C, and D are boolean random variables, and E is a random variable whose domain is $\{e_1, e_2, e_3, e_4, e_5\}$.

(i) How many entries are in the following probability tables and what is the sum of the values in each table? Write "?" if there is not enough information given. (1.0-point)

Table	Size	Sum
$P(e \mid B)$		
$P(A, B \mid c)$		
$P(A, B \mid C, d, E)$		
$P(a, E \mid B, C)$		
$P(A, c, E)$		

(ii) What is the **minimum** number of parameters needed to fully specify the distribution $P(A, B|C, d, E)$ (0.5-point)

(iii) What is the **minimum** number of parameters needed to fully specify the distribution $P(a, E|B, C)$ (0.5-point)

(b) Given the same set of random variables as defined in part (a). Write each of the following expressions in its simplest form, i.e., a single term. Make no independence assumptions unless otherwise stated.

Write "*Not possible*" if it is not possible to simplify the expression without making further independence assumptions.

(i)

$$\sum_{a'} P(a' \mid B, E) P(c \mid a', B, E)$$ (0.5-point)

(ii)

$$\frac{\sum_{a'} P(B \mid a', C) P(a' \mid C) P(C)}{\sum_{d', e'} P(d' \mid e', C) P(e' \mid C) P(C)}$$ (0.5-point)

Q2. Probability 2-point

(a) Select *all* of the expressions below that are equivalent to $P(A \mid B, C)$ given *no independence assumptions*. 0.5-point

- ◯ $\sum_d P(A \mid B, C, D = d)$
- ◯ $\sum_d P(A, D = d \mid B, C)$
- ◯ $P(A \mid B)P(A \mid C)$
- ◯ $P(A \mid C)$
- ◯ $P(A \mid B)P(B \mid C)$
- ◯ $P(A \mid C)P(C \mid B)$
- ◯ $\frac{P(A,B,C)}{P(B,C)}$
- ◯ $\frac{P(A)P(B|A)P(C|A,B)}{P(C)P(B|C)}$

(b) Select *all* of the expressions below that are equivalent to $P(A \mid B, C)$ given $A \perp\!\!\!\perp B$. 0.5-point

- ◯ $\sum_d P(A \mid B, C, D = d)$
- ◯ $\sum_d P(A, D = d \mid B, C)$
- ◯ $P(A \mid B)P(A \mid C)$
- ◯ $P(A \mid C)$
- ◯ $P(A \mid B)P(B \mid C)$
- ◯ $P(A \mid C)P(C \mid B)$
- ◯ $\frac{P(A,B,C)}{P(B,C)}$
- ◯ $\frac{P(A)P(B|A)P(C|A,B)}{P(C)P(B|C)}$

(c) Select *all* of the expressions below that are equivalent to $P(A \mid B, C)$ given $B \perp\!\!\!\perp C \mid A$. 0.5-point

- ◯ $\sum_d P(A \mid B, C, D = d)$
- ◯ $\sum_d P(A, D = d \mid B, C)$
- ◯ $P(A \mid B)P(A \mid C)$
- ◯ $P(A \mid C)$
- ◯ $P(A \mid B)P(B \mid C)$
- ◯ $P(A \mid C)P(C \mid B)$
- ◯ $\frac{P(A,B,C)}{P(B,C)}$
- ◯ $\frac{P(A)P(B|A)P(C|A,B)}{P(C)P(B|C)}$

(d) Select *all* of the expressions below that hold for any distribution over four random variables A, B, C and D. 0.5-point

- ◯ $P(A, B \mid C, D) = P(A \mid C, D)P(B \mid A, C, D)$
- ◯ $P(A, B) = P(A, B \mid C, D)P(C, D)$
- ◯ $P(A, B \mid C, D) = P(A, B)P(C, D)P(C, D \mid A, B)$
- ◯ $P(A, B \mid C, D) = P(A, B)P(D)P(C, D \mid A, B)$

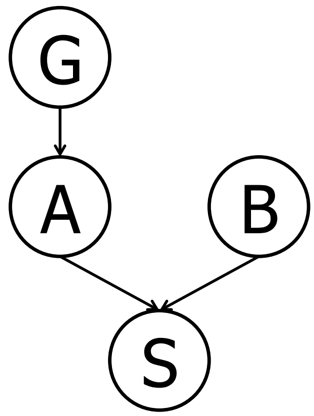
Q3. Bayes' Nets Representation and Probability

3-point

Suppose that a patient can have a symptom (S) that can be caused by two different diseases (A and B). It is known that the variation of gene G plays a big role in the manifestation of disease A. The Bayes' Net and corresponding conditional probability tables for this situation are shown below. For each part, you may leave your answer as an arithmetic expression.

$P(G)$	
$+g$	0.1
$-g$	0.9

$P(A \mid G)$		
$+g$	$+a$	1.0
$+g$	$-a$	0.0
$-g$	$+a$	0.1
$-g$	$-a$	0.9



$P(B)$	
$+b$	0.4
$-b$	0.6

$P(S \mid A, B)$			
$+a$	$+b$	$+s$	1.0
$+a$	$+b$	$-s$	0.0
$+a$	$-b$	$+s$	0.9
$+a$	$-b$	$-s$	0.1
$-a$	$+b$	$+s$	0.8
$-a$	$+b$	$-s$	0.2
$-a$	$-b$	$+s$	0.1
$-a$	$-b$	$-s$	0.9

(a) Compute the following entry from the joint distribution: 0.5-point

$P(+g, +a, +b, +s) =$

(b) What is the probability that a patient has disease A? 0.5-point

$P(+a) =$

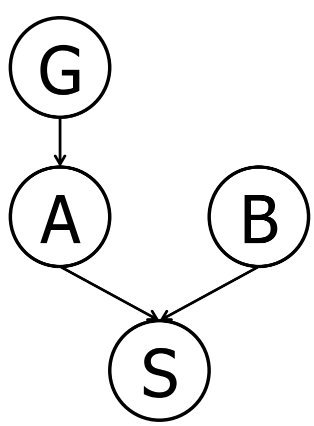
(c) What is the probability that a patient has disease A given that they have disease B? 0.5-point

$P(+a \mid +b) =$

The figures and table below are identical to the ones on the previous page and are repeated here for your convenience.

$P(G)$	
$+g$	0.1
$-g$	0.9

$P(A \mid G)$		
$+g$	$+a$	1.0
$+g$	$-a$	0.0
$-g$	$+a$	0.1
$-g$	$-a$	0.9



$P(B)$	
$+b$	0.4
$-b$	0.6

$P(S \mid A, B)$			
$+a$	$+b$	$+s$	1.0
$+a$	$+b$	$-s$	0.0
$+a$	$-b$	$+s$	0.9
$+a$	$-b$	$-s$	0.1
$-a$	$+b$	$+s$	0.8
$-a$	$+b$	$-s$	0.2
$-a$	$-b$	$+s$	0.1
$-a$	$-b$	$-s$	0.9

(d) What is the probability that a patient has disease A given that they have symptom S and disease B? 0.5-point

$P(+a|+s,+b) =$

(e) What is the probability that a patient has the disease carrying gene variation G given that they have disease A? 1-point

$P(+g|+a) =$