CS 188
Fall 2020

Regular Discussion 6 Solutions

Probability

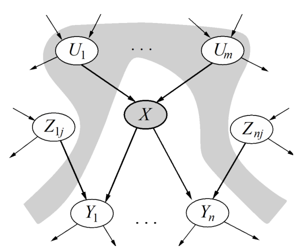
A **random variable** represents an event whose outcome is unknown. A **probability distribution** is an assignment of weights to outcomes A **joint distribution** over discrete random variables is a table of probabilities which captures the likelihood of each possible **outcome**, also known as an **assignment** of values to the random variables.

To write that random variables X and Y are **marginally independent**, we write $X \perp\!\!\!\perp Y$. To write that random variables X and Y are **conditionally independent** given another random variable Z, we write $X \perp\!\!\!\perp Y|Z$.

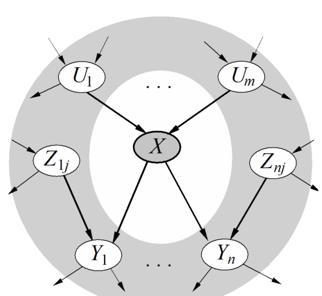
Bayesian Network Representation

In a Bayesian network, rather than storing information in a giant table, probabilities are instead distributed across a large number of smaller local probability tables along with a **directed acyclic graph** (DAG) which captures the relationships between variables. Thus, if we have a node representing variable X, we store $P(X|A_1, A_2, ..., A_N)$, where $A_1, ..., A_N$ are the parents of X.

- **Each node is conditionally independent of all its ancestor nodes (non-descendents) in the graph, given all of its parents**.

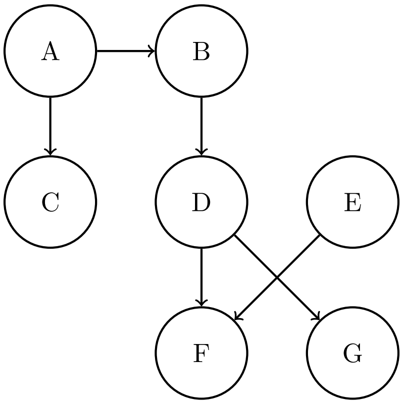


- **Each node is conditionally independent of all other variables given its Markov blanket.** A variable's Markov blanket consists of parents, children, children's other parents.



1 Bayes' Nets: Representation and Independence

Parts (a) and (b) pertain to the following Bayes' Net.



(a) Express the joint probability distribution as a product of terms representing individual conditional probabilities tables associated with the Bayes Net.

$P(A)P(C|A)P(B|A)P(D|B)P(E)P(F|D,E)P(G|D)$

(b) Assume each node can take on 4 values. How many entries do the factors at A, D, and F have?

A: 4

D: 4^2

F: 4^3

Consider the following probability distribution tables. The joint distribution $P(A, B, C, D)$ is equal to the product of these probability distribution tables.

A	$P(A)$
+a	0.8
–a	0.2

A	B	$P(B\|A)$
+a	+b	0.9
+a	–b	0.1
–a	+b	0.6
–a	–b	0.4

B	C	$P(C\|B)$
+b	+c	0.8
+b	–c	0.2
–b	+c	0.8
–b	–c	0.2

C	D	$P(D\|C)$
+c	+d	0.25
+c	–d	0.75
–c	+d	0.5
–c	–d	0.5

(c) State all non-conditional independence assumptions that are implied by the probability distribution tables.

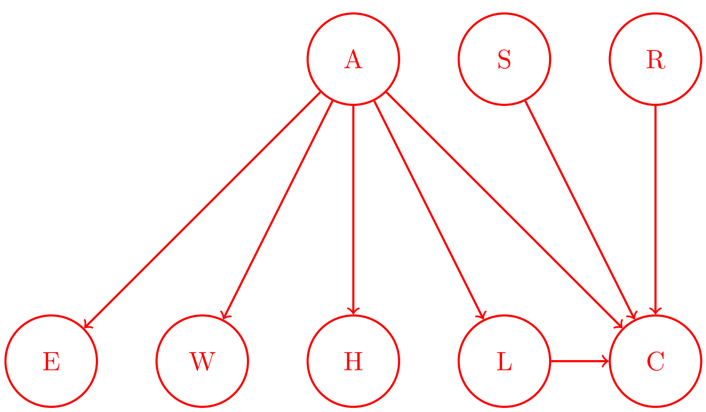
From the tables, we have $A \not\perp\!\!\!\perp B$ and $C \not\perp\!\!\!\perp D$. Then, we have every remaining pair of variables: $A \perp\!\!\!\perp C, A \perp\!\!\!\perp D, B \perp\!\!\!\perp C, B \perp\!\!\!\perp D$

You are building advanced safety features for cars that can warn a driver if they are falling asleep (A) and also calculate the probability of a crash (C) in real time. You have at your disposal 6 sensors (random variables):

- E: whether the driver's eyes are open or closed
- W: whether the steering wheel is being touched or not
- L: whether the car is in the lane or not
- S: whether the car is speeding or not
- H: whether the driver's heart rate is somewhat elevated or resting
- R: whether the car radar detects a close object or not

A influences $\{E, W, H, L, C\}$. C is influenced by $\{A, S, L, R\}$.

(d) Draw the Bayes Net associated with the description above by adding edges between the provided nodes where appropriate.

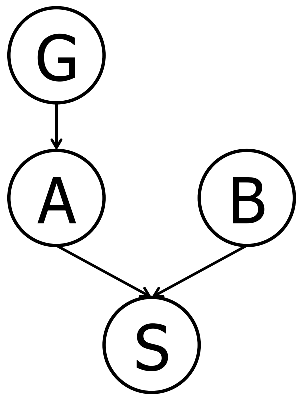


2 Bayes' Nets Representation and Probability

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

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



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

$P(S\|A,B)$			
$+a$	$+b$	$+s$	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:

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

$P(+g)P(+a|+g)P(+b)P(+s|+b,+a) = (0.1)(1.0)(0.4)(1.0) = 0.04$

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

$P(+a) = P(+a|+g)P(+g) + P(+a|-g)P(-g) = (1.0)(0.1) + (0.1)(0.9) = 0.19$

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

$P(+a|+b) = P(+a) = 0.19$ The first equality holds true as we have $A \perp\!\!\!\perp B$, which can be inferred from the graph of the Bayes' net.

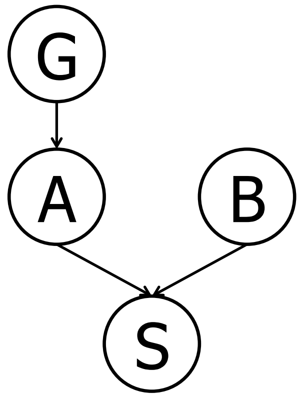
The figures and table below are identical to the ones on the previous page and are repeated here for your convenience.

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

$P(+a|+s,+b) = \frac{P(+a,+b,+s)}{P(+a,+b,+s)+P(-a,+b,+s)} = \frac{P(+a)P(+b)P(+s|+a,+b)}{P(+a)P(+b)P(+s|+a,+b)+P(-a)P(+b)P(+s|-a,+b)}$
$= \frac{(0.19)(0.4)(1.0)}{(0.19)(0.4)(1.0)+(0.81)(0.4)(0.8)} = \frac{0.076}{0.076+0.2592} \approx 0.2267$

$P(G)$	
$+g$	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

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

$$P(+g \mid +a) = \frac{P(+g)P(+a \mid +g)}{P(+g)P(+a \mid +g) + P(-g)P(+a \mid -g)} = \frac{(0.1)(1.0)}{(0.1)(1.0)+(0.9)(0.1)} = \frac{0.1}{0.1+0.09} = 0.5263$$