Probabilistic Reasoning

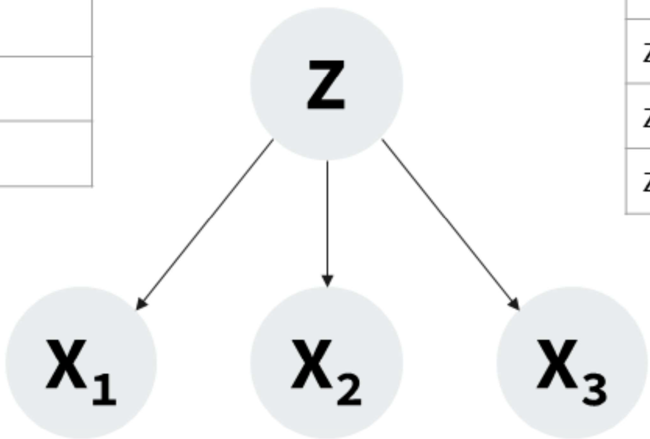
Chapter 14

14.1 We have a bag of three biased coins a, b, and c with probabilities of coming up heads of 20%, 60%, and 80%, respectively. One coin is drawn randomly from the bag (with equal likelihood of drawing each of the three coins), and then the coin is flipped three times to generate the outcomes X_1, X_2, and X_3.

a. Draw the Bayesian network corresponding to this setup and define the necessary CPTs.

b. Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

Z	P(Z)
Z = a	1/3
Z = b	1/3
Z = c	1/3



Z	P(X | Z)
Z = a	0.2
Z = b	0.6
Z = c	0.8

Assumptions:

1. Z is the type of the coin taken from the bag.
2. X is true if the flip resulted in a head, and false if tail.

B.

$P(Z \mid X_1, X_2, \neg X_3) = P(X_1, X_2, \neg X_3 | Z)\ P(Z) / P(X_1, X_2, \neg X_3 | Z) = \alpha\ P(X_1, X_2, \neg X_3 | Z)\ P(Z) =$
$\alpha\ P(X_1 | Z)\ P(X_2 | Z)\ P(\neg X_3 | Z)\ P(Z)$

$P(Z{=}a \mid X_1, X_2, \neg X_3) = \alpha * 0.2 * 0.2 * (1 - 0.2) * (1/3) = \alpha * 0.032/3$
$P(Z{=}b \mid X_1, X_2, \neg X_3) = \alpha * 0.6 * 0.6 * (1 - 0.6) * (1/3) = \alpha * 0.144/3$
$P(Z{=}c \mid X_1, X_2, \neg X_3) = \alpha * 0.8 * 0.8 * (1 - 0.8) * (1/3) = \alpha * 0.128/3$

So the coin is most likely to be of type “b”

If we want to calculate probabilities, we continue as follows:

Since $P(Z{=}a \mid X_1, X_2, \neg X_3) + P(Z{=}b \mid X_1, X_2, \neg X_3) + P(Z{=}c \mid X_1, X_2, \neg X_3) = 1$
So $\alpha = 1/(0.032/3 + 0.144/3 + 0.128/3) = 3/0.304$

$P(Z{=}a \mid X_1, X_2, \neg X_3) = 0.032/0.304 \cong 0.105$
$P(Z{=}b \mid X_1, X_2, \neg X_3) = 0.144/0.304 \cong 0.474$
$P(Z{=}c \mid X_1, X_2, \neg X_3) = 0.128/0.304 \cong 0.421$

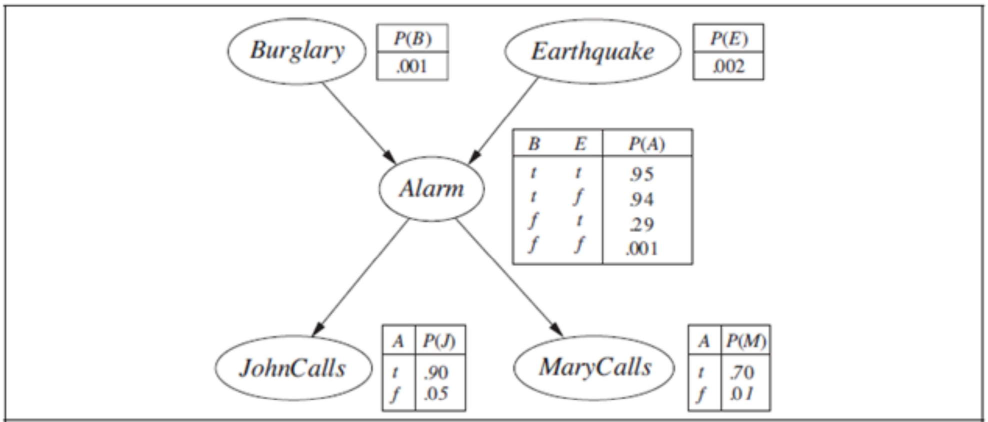


Figure 14.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters B, E, A, J, and M stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

14.4 Consider the Bayesian network in Figure 14.2.

a. If no evidence is observed, are *Burglary* and *Earthquake* independent? Prove this from the numerical semantics and from the topological semantics.

b. If we observe $Alarm = true$, are *Burglary* and *Earthquake* independent? Justify your answer by calculating whether the probabilities involved satisfy the definition of conditional independence.

A.

For the numerical proof, we would need to sample a lot of samples of the RVs and compute P(Burglary, Earthquake), P(Burglary) and P(Earthquake) then show that P(Burglary, Earthquake) = P(Burglary) * P(Earthquake) as the samples tend to infinity but that is infeasible so instead, we will emulate the process and compute the weights as follows.

First of all, we can sample from the network by sampler in a topological order:

1. Sample from “Burglary”. [P(Burglary)]
2. Sample from “Earthquake”. [P(Earthquake)]
3. Sample from “Alarm” given the sampled values of “Burglary” and “Earthquake”. [P(Alarm | Burglary, Earthquake)]
4. Sample from “JohnCalls” given the sampled value of “Alarm”. [P(JohnCalls | Alarm)]
5. Sample from “MaryCalls” given the sampled value of “Alarm”. [P(MaryCalls | Alarm)]

For the sake of simplicity, let us ignore “JohnCalls” and “MaryCalls” since they don’t depend on “Burglary” & “Earthquake” GIVEN “Alarm”.

So the probability of a certain tuple of values P(Burglary, Earthquake, Alarms) should be the product of the probability of each of value resulting from each sampling step (1 to 3) which will be: P(Burglary) * P(Earthquake) * P(Alarm | Burglary, Earthquake)

To compute P(Burglary, Earthquake) we need to sum P(Burglary, Earthquake, Alarm) over all possible values of “Alarm”, so

P(Burglary, Earthquake) = sum(P(Burglary) * P(Earthquake) * P(Alarm | Burglary, Earthquake)) over all possible values of “Alarm”
= P(Burglary) * P(Earthquake) * P(Alarm | Burglary, Earthquake) + P(Burglary) * P(Earthquake) * P(¬Alarm | Burglary, Earthquake)

By taking P(Burglary) * P(Earthquake) as a common factor, we end up with:

P(Burglary, Earthquake) = P(Burglary) * P(Earthquake) * (P(Alarm | Burglary, Earthquake) + P(¬Alarm | Burglary, Earthquake))

Since (P(Alarm | Burglary, Earthquake) + P(¬Alarm | Burglary, Earthquake)) = 1

P(Burglary, Earthquake) = P(Burglary) * P(Earthquake)

From the topological semantics, “Burglary” and “Earthquake” are d-separated so they are independent.

What is d-separation? This website does a great job at explaining it:
http://bayes.cs.ucla.edu/BOOK-2K/d-sep.html
To understand “d-separation”, we have to understand the concept of “collider”. A collider is node where the arrows along the path of interest face head-to-head. For example:

- In A -> B <- C, the node B is a collider for the path “A, B, C” since the arrows face head to head at it.
- In A -> B -> C, the node B is not a collider for the path “A, B, C”.
- In A <- B -> C, the node B is not a collider for the path “A, B, C”.

The d-separation rules are:

1. x and y are d-connected if there is an unblocked path (has no colliders) between them.
2. x and y are d-connected, conditioned on a set Z of nodes, if there is a collider-free path between x and y that traverses no member of Z. If no such path exists, we say that x and y are d-separated by Z, We also say then that every path between x and y is "blocked" by Z.
3. If a collider is a member of the conditioning set Z, or has a descendant in Z, then it no longer blocks any path that traces this collider.

Another way is that Given the parents of “Burglary” (there are none), Burglary is independent on any non-descendant (which includes “Earthquake”). Therefore “Burglary” and “Earthquake” are independent.

B.
No, Burglary and Earthquake are not independent given Alarm.

To show that they are dependant, we have to show that P(Burglary, Earthquake | Alarm) ≠ P(Burglary | Alarm) * P(Earthquake | Alarm) for any set of values.

P(Burglary, Earthquake | Alarm) = P(Alarm | Burglary, Earthquake) P(Burglary, Earthquake) / P(Alarm) = P(Alarm | Burglary, Earthquake) P(Burglary) P(Earthquake) /

P(Alarm)

P(Burglary, Earthquake | Alarm) = α P(Alarm | Burglary, Earthquake) P(Burglary) P(Earthquake)
P(Burglary, Earthquake | Alarm) = α 0.95 * 0.001 * 0.002 = α 0.0000019
P(Burglary, ¬Earthquake | Alarm) = α 0.94 * 0.001 * 0.998 = α 0.00093812
P(¬Burglary, Earthquake | Alarm) = α 0.29 * 0.999 * 0.002 = α 0.00057942
P(¬Burglary, ¬Earthquake | Alarm) = α 0.001 * 0.999 * 0.998 = α 0.000997002

α = 397.3864686728325

P(Burglary, Earthquake | Alarm) = 0.0007550342904783818
P(Burglary, ¬Earthquake | Alarm) = 0.3727961939913576
P(¬Burglary, Earthquake | Alarm) = 0.2302536676784126
P(¬Burglary, ¬Earthquake | Alarm) = 0.3961951040397514

P(Burglary | Alarm) = P(Burglary, Earthquake | Alarm) + P(Burglary, ¬Earthquake | Alarm) = 0.0007550342904783818 + 0.3727961939913576 = 0.373551228281836
P(Earthquake | Alarm) = P(Burglary, Earthquake | Alarm) + P(¬Burglary, Earthquake | Alarm) = 0.0007550342904783818 + 0.2302536676784126 = 0.231008701968891

P(Burglary | Alarm) P(Earthquake | Alarm) = 0.0862935843642718
P(Burglary, Earthquake | Alarm) = 0.0007550342904783818

So P(Burglary, Earthquake | Alarm) ≠ P(Burglary | Alarm) * P(Earthquake | Alarm)

14.6 Let H_x be a random variable denoting the handedness of an individual x, with possible values l or r. A common hypothesis is that left- or right-handedness is inherited by a simple mechanism; that is, perhaps there is a gene G_x, also with values l or r, and perhaps actual handedness turns out mostly the same (with some probability s) as the gene an individual possesses. Furthermore, perhaps the gene itself is equally likely to be inherited from either of an individual's parents, with a small nonzero probability m of a random mutation flipping the handedness.

a. Which of the three networks in Figure 14.20 claim that $\mathbf{P}(G_{father}, G_{mother}, G_{child}) = \mathbf{P}(G_{father})\mathbf{P}(G_{mother})\mathbf{P}(G_{child})$?

b. Which of the three networks make independence claims that are consistent with the hypothesis about the inheritance of handedness?

c. Which of the three networks is the best description of the hypothesis?

d. Write down the CPT for the G_{child} node in network (a), in terms of s and m.

e. Suppose that $P(G_{father} = l) = P(G_{mother} = l) = q$. In network (a), derive an expression for $P(G_{child} = l)$ in terms of m and q only, by conditioning on its parent nodes.

f. Under conditions of genetic equilibrium, we expect the distribution of genes to be the same across generations. Use this to calculate the value of q, and, given what you know about handedness in humans, explain why the hypothesis described at the beginning of this question must be wrong.

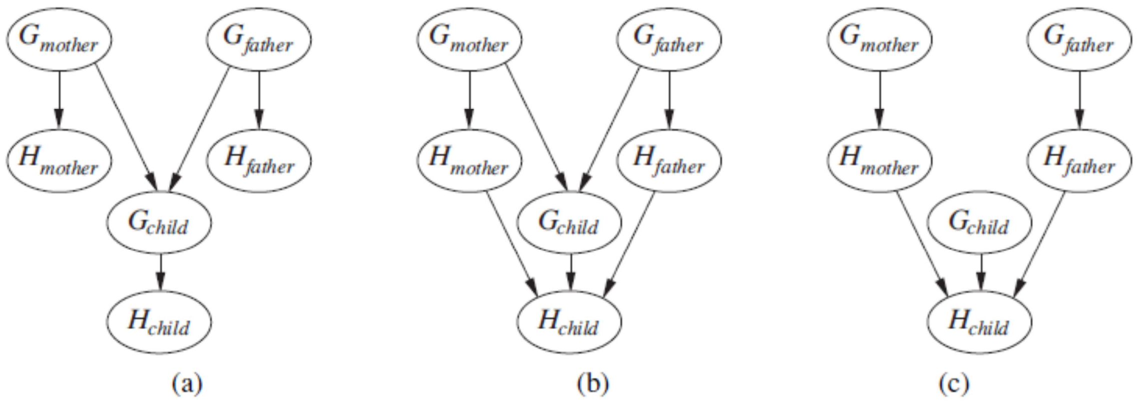


Figure 14.20 Three possible structures for a Bayesian network describing genetic inheritance of handedness.

A.
Figure (c).

B.
Figure (a) and (b). There are extra unneeded arcs but they don't violate the conditional dependence.

C.
Figure (a) since it represents the dependencies correctly with less arcs (more efficient).

D.
Assuming Gc, Gm, Gf means the the child, the mother and the father are left-handed respectively.
P(Gc | Gm, Gf) = 1 - m
P(Gc | Gm, ¬Gf) = 0.5
P(Gc | ¬Gm, Gf) = 0.5
P(Gc | ¬Gm, ¬Gf) = m

Assuming H means the individual is left-handed.
P(H | G) = s
P(H | ¬G) = 1 - s

E.
P(Gc) = sum P(Gc | Gm, Gf)P(Gm)P(Gf) over all values of Gm, Gf =
P(Gc) = (1-m)*q*q + 0.5*q*(1-q) + 0.5*(1-q)*q + m*(1-q)*(1-q)
        = q^2 - m*q^2 + q - q^2 + m - m*2*q + m*q^2

$= q + m - 2mq$

F.
Equilibrium means the P(G) should be the same across generations so
$P(Gc) = P(Gm) = P(Gf) = q$
$q + m - 2mq = q$
$m = 2mq$

Since $m > 0$,
$1 = 2q$

So $q = 0.5$

Which means that there is a 50% that any person is left-handed and 50% that they are right-handed.
This is not true in reality so we can infer that the hypothesis proposed by the question is wrong.

14.11 In your local nuclear power station, there is an alarm that senses when a temperature gauge exceeds a given threshold. The gauge measures the temperature of the core. Consider the Boolean variables A (alarm sounds), F_A (alarm is faulty), and F_G (gauge is faulty) and the multivalued nodes G (gauge reading) and T (actual core temperature).

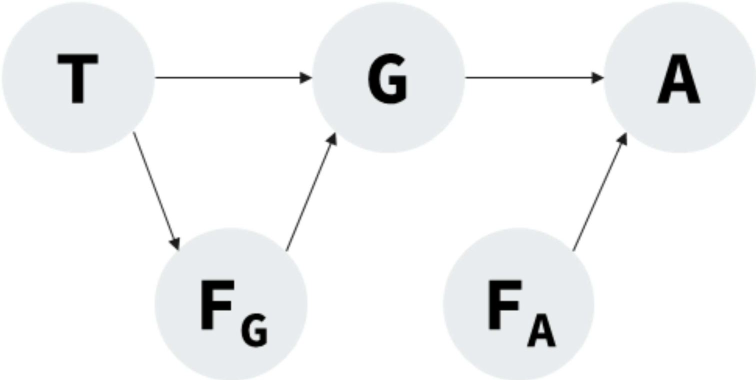
a. Draw a Bayesian network for this domain, given that the gauge is more likely to fail when the core temperature gets too high.

b. Is your network a polytree? Why or why not?

c. Suppose there are just two possible actual and measured temperatures, normal and high; the probability that the gauge gives the correct temperature is x when it is working, but y when it is faulty. Give the conditional probability table associated with G.

d. Suppose the alarm works correctly unless it is faulty, in which case it never sounds. Give the conditional probability table associated with A.

e. Suppose the alarm and gauge are working and the alarm sounds. Calculate an expression for the probability that the temperature of the core is too high, in terms of the various conditional probabilities in the network.



T	F_G	$P(G = normal \mid T, F_G)$
normal	Ok	x
normal	Faulty	y
high	Ok	1-x
high	Faulty	1-y

G	F_A	$P(A = sounds \mid G, F_A)$
normal	Ok	0
normal	Faulty	0
high	Ok	1
high	Faulty	0

B.
No, it is not a poly tree, since there are 2 paths from T to G.

E.
Let t mean T = High
Let g mean G = High
Let f_G mean F_G = Faulty
Let f_A mean F_A = Faulty
Let a mean A = sounds

$P(t \mid a, \neg f_G, \neg f_A)$ = sum [$P(t \mid G, a, \neg f_G, \neg f_A)\, P(G \mid a, \neg f_G, \neg f_A)$] over G
= sum [$P(t \mid G, \neg f_G)\, P(G \mid a, \neg f_A)$] over G

For the first term:
$P(t \mid G, \neg t_G) = P(G, t, \neg f_G) / P(G, \neg f_G)$
and $P(G, t, \neg f_G) = P(G \mid t, \neg f_G)\, P(t, \neg f_G)$

So
$P(g, t, \neg f_G) = x * P(t, \neg f_G)$
$P(\neg g, t, \neg f_G) = (1-x) * P(t, \neg f_G)$
$P(g, \neg t, \neg f_G) = (1-x) * P(\neg t, \neg f_G)$
$P(\neg g, \neg t, \neg f_G) = x * P(\neg t, \neg f_G)$

$P(g, \neg f_G) = x * P(t, \neg f_G) + (1-x) * P(\neg t, \neg f_G)$
$P(\neg g, \neg f_G) = (1-x) * P(t, \neg f_G) + x * P(\neg t, \neg f_G)$

$P(t \mid G, \neg t_G) = \langle P(g, t, \neg f_G), P(\neg g, t, \neg f_G)\rangle / \langle P(g, \neg f_G), P(\neg g, \neg f_G)\rangle$
$= \langle x * P(t, \neg f_G), (1-x) * P(t, \neg f_G)\rangle / \langle x * P(t, \neg f_G) + (1-x) * P(\neg t, \neg f_G), (1-x) * P(t, \neg f_G) + x * P(\neg t, \neg f_G)\rangle$

For the 2nd term:
$P(G \mid a, \neg f_A) = P(a \mid G, \neg f_A) P(G) P(\neg f_A) / P(a, \neg f_A) = \langle P(a \mid g, \neg f_A) P(g) P(\neg f_A) / P(a, \neg f_A), P(a \mid \neg g, \neg f_A) P(\neg g) P(\neg f_A) / P(a, \neg f_A)\rangle = \langle 1, 0\rangle$

Therefore
$P(t \mid a, \neg f_G, \neg f_A) = P(t \mid g, \neg f_G) = x * P(t, \neg f_G) / (x * P(t, \neg f_G) + (1-x) * P(\neg t, \neg f_G))$

Since $P(t, \neg f_G) = P(\neg f_G \mid t) P(t)$ and $P(\neg t, \neg f_G) = P(\neg f_G \mid \neg t) P(\neg t)$

$P(t \mid a, \neg f_G, \neg f_A) = x * P(\neg f_G \mid t) P(t) / (x * P(\neg f_G \mid t) P(t) + (1-x) * P(\neg f_G \mid \neg t) P(\neg t))$

14.14 Consider the Bayes net shown in Figure 14.23.

a. Which of the following are asserted by the network *structure*?

(i) $\mathbf{P}(B, I, M) = \mathbf{P}(B)\mathbf{P}(I)\mathbf{P}(M)$.

(ii) $\mathbf{P}(J \mid G) = \mathbf{P}(J \mid G, I)$.

(iii) $\mathbf{P}(M \mid G, B, I) = \mathbf{P}(M \mid G, B, I, J)$.

b. Calculate the value of $P(b, i, \neg m, g, j)$.

c. Calculate the probability that someone goes to jail given that they broke the law, have been indicted, and face a politically motivated prosecutor.

d. A **context-specific independence** (see page 542) allows a variable to be independent of some of its parents given certain values of others. In addition to the usual conditional independences given by the graph structure, what context-specific independences exist in the Bayes net in Figure 14.23?

e. Suppose we want to add the variable $P = PresidentialPardon$ to the network; draw the new network and briefly explain any links you add.

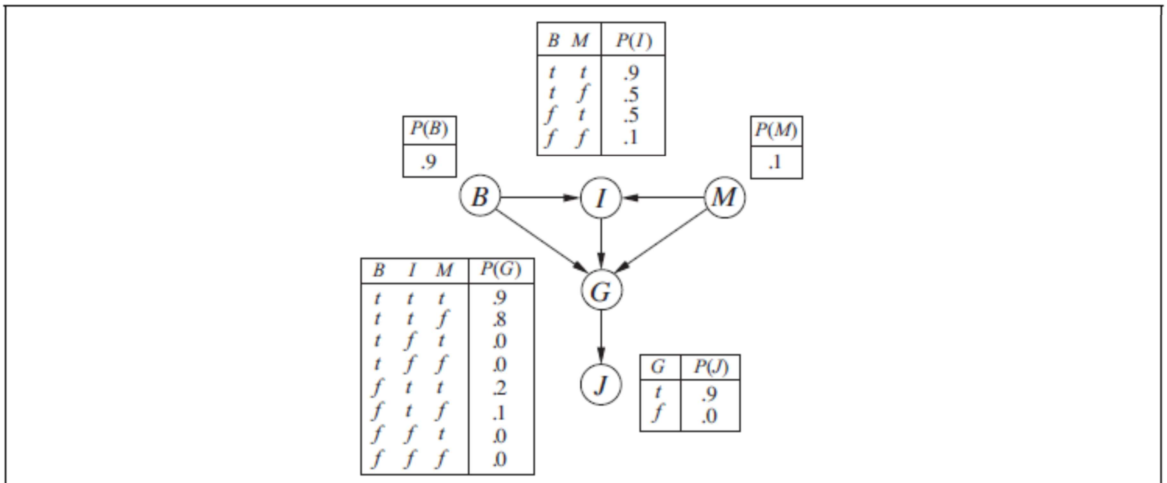


Figure 14.23 A simple Bayes net with Boolean variables $B = BrokeElectionLaw$, $I = Indicted$, $M = PoliticallyMotivatedProsecutor$, $G = FoundGuilty$, $J = Jailed$.

A.
(i) No. B, I, M are not independent.
(ii) Yes, J is independent of I given G.
(iii) Yes. Since G is given, it blocks all the paths from M to J so they are d-separated.

B.
P(b, i, ¬m, g, j) = P(j | g) P(g | b, i, ¬m) P(i | b, ¬m) P(b) P(¬m) = 0.9 * 0.8 * 0.5 * 0.9 * (1 - 0.1)

C.
P(j | b, i, m) = P(j | g)P(g | b, i, m) + P(j | ¬g)P(¬g | b, i, m) = 0.9 * 0.9 + 0 * 0.1 = 0.81

D.
G is independent on B and M given I=False since P(G | B, ¬i, M) = 0 = P(G | ¬i)

E.
A pardon is unnecessary if the person is not indicted or not found guilty; so I and G are parents of P. One could also add B and M as parents of P, since a pardon is more likely if the person is actually innocent and if the prosecutor is politically motivated. (There are other causes of Pardon, such as LargeDonationToPresidentsParty, but such variables are not currently in the model.) The pardon (presumably) is a getout- of-jail-free card, so P is a parent of J.