14.5 Suppose that in a Bayesian network containing an unobserved variable Y, all the variables in the Markov blanket MB(Y) have been observed.
a. Prove that removing the node Y from the network will not affect the posterior distribution for any other unobserved variable in the network.
b. Discuss whether we can remove Y if we are planning to use (i) rejection sampling and (ii) likelihood weighting.

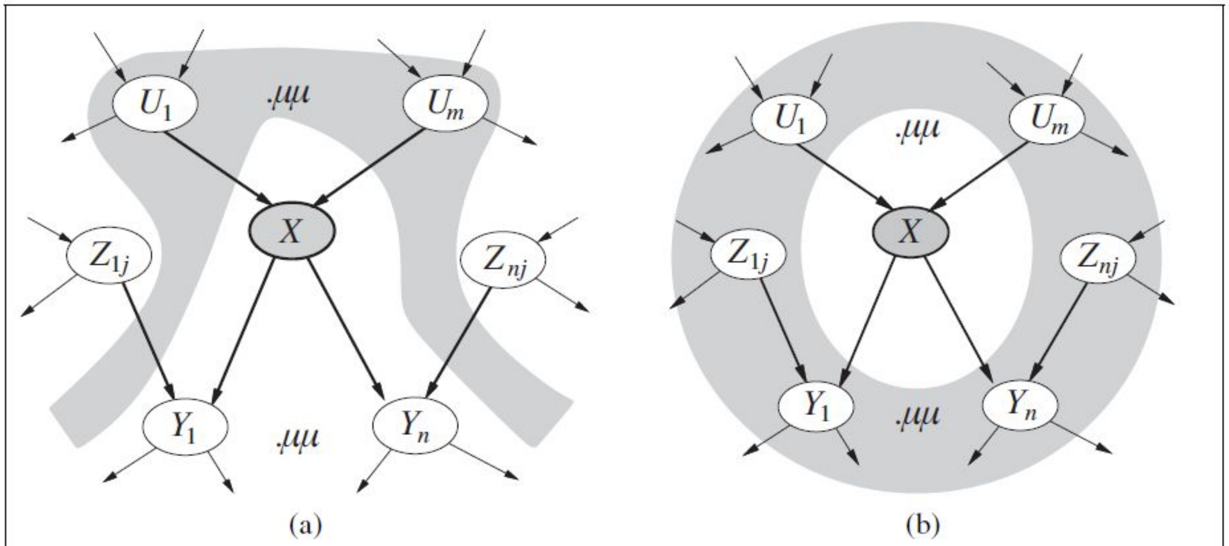

Figure 14.4 (a) A node X is conditionally independent of its non-descendants (e.g., the Z_{ij}s) given its parents (the U_is shown in the gray area). (b) A node X is conditionally independent of all other nodes in the network given its Markov blanket (the gray area).

a. Prove that removing the node Y from the network will not affect the posterior distribution for any other unobserved variable in the network.

Extracted from the book: *"Another important independence property is implied by the topological semantics: a node is conditionally independent of all other nodes in the network, given its parents, children, and children's parents—that is, given its Markov blanket."* [AIMA 3E, p.517]]

Let's call:

$$MB(Y) = \{ Parents(Y);\ Children(Y);\ Parents(Children(Y)) \}$$
$$mb(Y) = observed\ values\ for\ each\ variable\ that\ belongs\ to\ MB(Y)$$

We can try to resolve the prior probability of an unobservable variable X_i that is different from variable Y and that at the same time X_i does not belong to the Markov Blanket of Y given $mb(Y)$ and Y . By the independence property mentioned above, If we know that X_i does not depends on Y , then Y does not provide information about X_i and because of this we can say that:

$$P(X_i \mid Y \wedge mb(Y)) = P(X_i \mid mb(Y))$$

Let's look how is that

I) By independence definition between X_i and Y given $mb(Y)$:

$$P(X_i, Y \mid mb(Y)) = P(X_i \mid mb(Y)) \times P(Y \mid mb(Y))$$

II) By conditional definition:

$$P(X_i, Y \mid mb(Y)) = P(X_i,\ Y,\ mb(Y)) / P(mb(Y))$$

III) Matching (I) and (II)

$$P(X_i \mid mb(Y)) \times P(Y \mid mb(Y)) = P(X_i,\ Y,\ mb(Y)) / P(mb(Y))$$

IV) Using condition definition for $P(Y \mid mb(Y))$ in (III)

$$P(X_i \mid mb(Y)) \times P(Y, mb(Y)) \,/\, P(mb(Y)) = P(X_i,\ Y,\ mb(Y)) / P(mb(Y))$$

V) Reducing (IV):

$$P(X_i \mid mb(Y)) \times P(Y, mb(Y)) = P(X_i,\ Y,\ mb(Y))$$

VI) Using product rule in right hand of (IV):

$$P(X_i \mid mb(Y)) \times P(Y, mb(Y)) = P(X_i \mid Y,\ mb(Y)) \times P(Y,\ mb(Y))$$

VII) Reducing (VI):

$$P(X_i \mid mb(Y)) = P(X_i \mid Y,\ mb(Y))$$

Q.E.D.

Another important question is: what does it happen with the children of Y when we remove node Y? Probably children nodes of Y will be linked to the original parents of Y, and the CPTs for variables in $Children(Y)$ will require to be updated (this is a new Bayesian Network).

b. Discuss whether we can remove Y if we are planning to use (i) rejection sampling and (ii) likelihood weighting.

(i) If children of Y are present in the Bayesian Network, then Rejection Sampling could be used if we recompute all the CPTs mentioned above. These are the CPTs of the variables in children(Y).

(ii) The likelihood weighting could be used too but we will require to update the CPTs of the nodes that belong to original children nodes of Y.

Conclusion: both sampling algorithms can be used but it is important to note that if we remove a node with children from a Bayesian Network, then it will be required to update children nodes with new possible parents and then recalculate the CPTs of these children. But if we are interested in variables outside of the Markov Blanket of Y, then we could use the original CPTs in the Bayesian Network.