14.2 Equation (14.1) on page 513 defines the joint distribution represented by a Bayesian network in terms of the parameters θ(Xi |Parents(Xi)). This exercise asks you to derive the equivalence between the parameters and the conditional probabilities P(Xi |Parents(Xi)) from this definition.
a. Consider a simple network X → Y → Z with three Boolean variables. Use Equations (13.3) and (13.6) (pages 485 and 492) to express the conditional probability P(z | y) as the ratio of two sums, each over entries in the joint distribution P(X, Y, Z).
b. Now use Equation (14.1) to write this expression in terms of the network parameters θ(X), θ(Y |X), and θ(Z | Y).
c. Next, expand out the summations in your expression from part (b), writing out explicitly the terms for the true and false values of each summed variable. Assuming that all network parameters satisfy the constraint xi θ(xi | parents(Xi))=1, show that the resulting expression reduces to θ(x | y).
d. Generalize this derivation to show that θ(Xi |Parents(Xi)) = P(Xi |Parents(Xi)) for any Bayesian network.

a. Consider a simple network X → Y → Z with three Boolean variables. Use Equations (13.3) and (13.6) (pages 485 and 492) to express the conditional probability P(z | y) as the ratio of two sums, each over entries in the joint distribution P(X, Y, Z).

Equation 13.3 (Conditional Probability defined in terms of unconditional probabilities)

$$P(a \mid b) = P(a \wedge b) \,/\, P(b)$$

Equation 13.6 (Marginalization Rule)

$$P(Y) = \sum_{z \in Z} P(Y \wedge z)$$

I) By equation 13.3:

$$P(z \mid y) = P(z \wedge y) \,/\, P(y) = P(y \wedge z) \,/\, P(y)$$

II) By equation 13.6

$$P(y \wedge z) = \sum_{x \in X} P(y \wedge z \wedge x) = \sum_{x \in X} P(x \wedge y \wedge z)$$

III) By equation 13.6:

$$P(y) = \sum_{x \in X, z \in Z} P(y \wedge x \wedge z) = \sum_{x \in X, z \in Z} P(x \wedge y \wedge z)$$

IV) By (I) (II) and (III)

$$P(z \mid y) = \Big(\sum_{x \in X} P(x \wedge y \wedge z)\Big) \,/\, \sum_{x \in X, z \in Z} P(x \wedge y \wedge z)$$

Q.E.D.

b. Now use Equation (14.1) to write this expression in terms of the network parameters θ(X), θ(Y |X), and

θ(Z | Y).

Equation 14.1

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} \theta(x_i \mid parents(X_i))$$

I) By equation 14.1

$$P(z \mid y) = \frac{\sum_{x \in X} \theta(x \mid parents(x)) \times \theta(y \mid parents(y)) \times \theta(z \mid parents(z))}{\sum_{x \in X, z \in Z} \theta(x \mid parents(x)) \times \theta(y \mid parents(y)) \times \theta(z \mid parents(z))}$$

$$P(z \mid y) = \frac{\sum_{x \in X} \theta(x) \times \theta(y \mid x) \times \theta(z \mid y)}{\sum_{x \in X, z \in Z} \theta(x) \times \theta(y \mid x) \times \theta(z \mid y)}$$

c. Next, expand out the summations in your expression from part (b), writing out explicitly the terms for the true and false values of each summed variable. Assuming that all network parameters satisfy the constraint xi θ(xi | parents(Xi))=1, show that the resulting expression reduces to θ(x | y).

$$P(z \mid y) = \frac{\theta(x) \times \theta(y \mid x) \times \theta(z \mid y) + \theta(-x) \times \theta(y \mid -x) \times \theta(z \mid y)}{\theta(x) \times \theta(y \mid x) \times \theta(z \mid y) + \theta(-x) \times \theta(y \mid -x) \times \theta(z \mid y) + \theta(x) \times \theta(y \mid x) \times \theta(-z \mid y) + \theta(-x) \times \theta(y \mid -x) \times \theta(-z \mid y)}$$

$$P(z \mid y) = \frac{\theta(z \mid y) \times (\theta(x) \times \theta(y \mid x) + \theta(-x) \times \theta(y \mid -x))}{\theta(z \mid y) \times [\theta(x) \times \theta(y \mid x) + \theta(-x) \times \theta(y \mid -x)] + \theta(-z \mid y) \times [\theta(x) \times \theta(y \mid x) + \theta(-x) \times \theta(y \mid -x)]}$$

$$P(z \mid y) = \frac{\theta(z \mid y) \times (\theta(x) \times \theta(y \mid x) + \theta(-x) \times \theta(y \mid -x))}{[\theta(z \mid y) + \theta(-z \mid y)] \times [\theta(x) \times \theta(y \mid x) + \theta(-x) \times \theta(y \mid -x)]}$$

$$P(z \mid y) = \frac{\theta(z \mid y) \times (\theta(x) \times \theta(y \mid x) + \theta(-x) \times \theta(y \mid -x))}{1 \times [\theta(x) \times \theta(y \mid x) + \theta(-x) \times \theta(y \mid -x)]}$$

$$P(z \mid y) = \theta(z \mid y)$$

Q.E.D.

Another simpler way:

$$P(z \mid y) = \frac{\sum_{x \in X} \theta(x) \times \theta(y \mid x) \times \theta(z \mid y)}{\sum_{x \in X, z \in Z} \theta(x) \times \theta(y \mid x) \times \theta(z \mid y)} = \frac{\theta(z \mid y) \times \sum_{x \in X} \theta(x) \times \theta(y \mid x)}{\sum_{x \in X} \theta(x) \times \theta(y \mid x) \ \sum_{z \in Z} \theta(z \mid y)}$$

$$P(z \mid y) = \frac{\theta(z \mid y)}{\sum_{z \in Z} \theta(z \mid y)} = \frac{\theta(z \mid y)}{1} = \theta(z \mid y)$$

d. Generalize this derivation to show that θ(Xi |Parents(Xi)) = P(Xi |Parents(Xi)) for any Bayesian network.

$Domain\ of\ variables\ in\ the\ Bayesian\ Network\ X = \{X_1,\ X_2,\ ...,\ X_n\}$

I) By conditional definition:

$P(X_i \mid parents(X_i)) = P(X_i \wedge parents(X_i)) \,/\, P(parents(X_i))$

II) By marginalization rule:

$$P(X_i \wedge parents(X_i)) = \sum_{z\, \in\, Z} P(X_i \wedge parents(X_i) \wedge z)$$

III) By marginalization rule:

$$P(parents(X_i)) = \sum_{z\, \in\, Z} P(parents(X_i) \wedge z)$$

(*) The above z represents joint value of the form $(x_1, x_2, ..., x_j, ...,\ x_n)$ and Z is the full joints world.

IV) Replacing (II) and (III) in equation (I):

$$P(X_i \mid parents(X_i)) = \sum_{z\, \in\, Z} P(X_i \wedge parents(X_i) \wedge z) \,/ \sum_{z\, \in\, Z} P(parents(X_i) \wedge z)$$

V) By equation 14.1

$$P(z_1,\ z_2,\ ...,\ z_n) = \prod_{i=1}^{n} \theta(z_i \mid parents(Z_i))$$

(*) The above z represents a joint of the form $(x_1, x_2, ..., x_j, ...,\ x_n)$ and z_i represents the literal value of the variable $Z_i = X_i$.

VI) By equation 14.1

$$P(parents(X_i),\ x_i,\ z) = \theta(x_i \mid parents(X_i) \times \prod_{p\, \in\, P,\, P \in\, parents(X_i)} \theta(p \mid parents(P) \times \prod_{j=1}^{m} \theta(z_j \mid parents(Z_j))$$

VII) By equation 14.1

$$P(parents(X_i),\ z) = \prod_{p\, \in\, P,\, P \in\, parents(X_i)} \theta(p \mid parents(P) \times \prod_{j=1}^{m} \theta(z_j \mid parents(Z_j))$$

VIII) Replacing (VI) and (VII) in (IV):

$$P(X_i \mid parents(X_i)) = \frac{\sum_{z\, \in\, Z} \theta(x_i \mid parents(X_i) \times \prod_{p\, \in\, P,\, P \in\, parents(X_i)} \theta(p \mid parents(P) \times \prod_{j=1}^{m} \theta(z_j \mid parents(Z_j))}{\sum_{z\, \in\, Z} \prod_{p\, \in\, P,\, P \in\, parents(X_i)} \theta(p \mid parents(P) \times \prod_{j=1}^{m} \theta(z_j \mid parents(Z_j))}$$

$$P(X_i \mid parents(X_i)) = \frac{\theta(x_i \mid parents(X_i) \times \sum_{z\, \in\, Z} \prod_{p\, \in\, P,\, P \in\, parents(X_i)} \theta(p \mid parents(P) \times \prod_{j=1}^{m} \theta(z_j \mid parents(Z_j))}{\sum_{z\, \in\, Z} \prod_{p\, \in\, P,\, P \in\, parents(X_i)} \theta(p \mid parents(P) \times \prod_{j=1}^{m} \theta(z_j \mid parents(Z_j))}$$

$$P(X_i \mid parents(X_i)) = \theta(X_i \mid parents(X_i))$$

Q.E.D.