3.1

Exercise 3.1

Provide an example of a distribution $P(X_1, X_2, X_3)$ where for each $i \neq j$, we have that $(X_i \perp X_j) \in \mathcal{I}(P)$, but we also have that $(X_1, X_2 \perp X_3) \notin \mathcal{I}(P)$.

$X_1 \sim \text{Bernoulli}(0.5)$

$X_2 \sim \text{Bernoulli}(0.5)$

$$X_3 = \begin{cases} 1 & \text{if } X_1 = X_2 \\ 0 & \text{else} \end{cases}$$

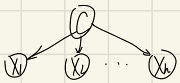
3.2

Exercise 3.2

a. Show that the naive Bayes factorization of equation (3.7) follows from the naive Bayes independence assumptions of equation (3.6).

b. Show that equation (3.8) follows from equation (3.7).

c. Show that, if all the variables $C, X_1, \ldots, X_n$ are binary-valued, then $\log \frac{P(C=c^1|x_1,\ldots,x_n)}{P(C=c^2|x_1,\ldots,x_n)}$ is a linear function of the value of the finding variables, that is, can be written as $\sum_{i=1}^n \alpha_i X_i + \alpha_0$ (where $X_i = 0$ if $X = x^0$ and 1 otherwise).



$$p(C, X_1, \ldots, X_n) = P(C) \cdot p(X_1|C) \cdot p(X_2|X_1, C) \cdots p(X_n|X_{n-1} \ldots X_1\, C)$$

$$= P(C) \cdot p(X_1|C) \cdot p(X_2|C) \cdots p(X_n|C)$$

$$= P(C) \cdot \prod_{i=1}^{n} P(X_i|C)$$

$$\frac{P(C=C^1|X_1 \ldots X_n)}{P(C=C^2|X_1 \ldots X_n)} = \frac{P(C=C^1, X_1 \ldots X_n)/P(X_1, \ldots X_n)}{P(C=C^2, X_1 \ldots X_n)/P(X_1 \ldots X_n)}$$

$$= \frac{P(C=C^1) \cdot P(X_1|C=C^1)\ P(X_2|X_1, C=C^1) \cdots}{P(C=C^2) \cdot P(X_1|C=C^2) \cdot P(X_2|X_1, C=C^2) \cdots}$$

$$= \frac{P(C=C^1)}{P(C=C^2)} \cdot \prod_{i=1}^{n} \frac{P(X_i|C=C^1)}{P(X_i|C=C^2)}$$

	X_i^0	X_i^1
$C=C^1$	$1-p_i$	p_i
$C=C^2$	$1-q_i$	q_i

$$\log \frac{P(C=C^1|X_1 \ldots X_n)}{P(C=C^2|X_1 \ldots X_n)} = \log \frac{P(C=C^1)}{P(C=C^2)} + \sum_{i=1}^{n} \log \frac{P(X_i|C=C^1)}{P(X_i|C=C^2)}$$

$$= \left[\log P(C=C^1) - \log P(C=C^2)\right] + \sum_{i=1}^{n} \left[\log P(X_i|C=C^1) - \log P(X_i|C=C^2)\right]$$

$$= \left[\log P(C=C^1) - \log P(C=C^2)\right] + \sum_{i=1}^{n} \left[\log p_i^{X_i}(1-p_i)^{(1-X_i)} - \log q_i^{X_i}(1-q_i)^{(1-X_i)}\right]$$

$$= \cdots + \sum_{i=1}^{n} \left[X_i \log p_i + (1-X_i)\log(1-p_i) - X_i \log q_i - (1-X_i)\log(1-q_i)\right]$$

$$= \alpha_0 + \sum_{i=1}^{n} \alpha_i X_i$$

3.3.

Exercise 3.3

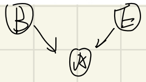
Consider a simple example (due to Pearl), where a burglar alarm (A) can be set off by either a burglary (B) or an earthquake (E).

a. Define constraints on the CPD of $P(A \mid B, E)$ that imply the *explaining away* property.

b. Show that if our model is such that the alarm always (deterministically) goes off whenever there is a earthquake:

$$P(a^1 \mid b^1, e^1) = P(a^1 \mid b^0, e^1) = 1$$

then $P(b^1 \mid a^1, e^1) = P(b^1)$, that is, observing an earthquake provides a full explanation for the alarm.



$$p(b^1 \mid a^1 e^1) = \frac{P(b^1 a^1 e^1)}{P(a^1 e^1)} = \frac{P(a^1 \mid b^1 e^1) \cdot p(b^1 e^1)}{p(a^1 e^1)}$$

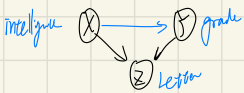
$$= \frac{p(a^1 \mid e^1)\ p(b^1 e^1)}{p(a^1 e^1)} = \frac{p(b^1 e^1)}{p(e^1)} = p(b^1)$$

3.4

Exercise 3.4

We have mentioned that explaining away is one type of intercausal reasoning, but that other type of intercausal interactions are also possible. Provide a realistic example that exhibits the opposite type of interaction. More precisely, consider a v-structure $X \to Z \leftarrow Y$ over three binary-valued variables. Construct a CPD $P(Z \mid X, Y)$ such that:

- X and Y both increase the probability of the effect, that is, $P(z^1 \mid x^1) > P(z^1)$ and $P(z^1 \mid y^1) > P(z^1)$,
- each of X and Y increases the probability of the other, that is, $P(x^1 \mid z^1) < P(x^1 \mid y^1, z^1)$, and similarly $P(y^1 \mid z^1) < P(y^1 \mid x^1, z^1)$.



x^1 = high intelligence

y^1 = high grade

z^1 = strong recommendation

3.5

Exercise 3.5

Consider the Bayesian network of figure 3.14.

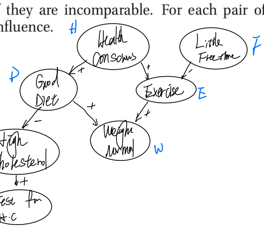
Assume that all variables are binary-valued. We do not know the CPDs, but do know how each random variable *qualitatively* affects its children. The influences, shown in the figure, have the following interpretation:

- $X \overset{+}{\to} Y$ means $P(y^1 \mid x^1, \boldsymbol{u}) > P(y^1 \mid x^0, \boldsymbol{u})$, for all values $\boldsymbol{u}$ of Y's other parents.
- $X \overset{-}{\to} Y$ means $P(y^1 \mid x^1, \boldsymbol{u}) < P(y^1 \mid x^0, \boldsymbol{u})$, for all values $\boldsymbol{u}$ of Y's other parents.

We also assume explaining away as the interaction for all cases of intercausal reasoning.

For each of the following pairs of conditional probability queries, use the information in the network to determine if one is larger than the other, if they are equal, or if they are incomparable. For each pair of queries, indicate all relevant active trails, and their direction of influence.

(a)	$P(t^1 \mid d^1)$	$P(t^1)$
(b)	$P(d^1 \mid t^0)$	$P(d^1)$
(c)	$P(h^1 \mid e^1, f^1)$	$P(h^1 \mid e^1)$
(d)	$P(c^1 \mid f^0)$	$P(c^1)$
(e)	$P(c^1 \mid h^0)$	$P(c^1)$
(f)	$P(c^1 \mid h^0, f^0)$	$P(c^1 \mid h^0)$
(g)	$P(d^1 \mid h^1, e^0)$	$P(d^1 \mid h^1)$
(h)	$P(d^1 \mid e^1, f^0, w^1)$	$P(d^1 \mid e^1, f^0)$
(i)	$P(t^1 \mid w^1, f^0)$	$P(t^1 \mid w^1)$



(a) $p(t^1 \mid d^1) < p(t^1)$ $\quad D \to C \to T$

(b) $p(d^1 \mid t^0) > p(d^1)$ $\quad D \to C \to T$

(c) $p(h^1 \mid e^1, f^1) > p(h^1 \mid e^1)$ $\quad H \to E \leftarrow F$

(d). $p(c^1 \mid f^0) = p(c^1)$

(e). $p(c^1 \mid h^0) > p(c^1)$ $\quad H \to D \to$

(f). $p(c^1 \mid h^0, f^0) = p(c^1 \mid h^0)$

(g) $p(d^1 \mid h^1 e^0) = p(d^1 \mid h^1)$

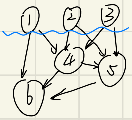
3.6 $X_1, \dots X_n$ X_i has $|Val(X_i)| = l$

a. Assume that we have a Bayesian network over $X_1, \dots, X_n$, such that each node has at most k parents. What is a simple upper bound on the number of independent parameters in the Bayesian network? How many independent parameters are in the full joint distribution over $X_1, \dots, X_n$?

b. Now, assume that each variable X_i has the parents $X_1, \dots, X_{i-1}$. How many independent parameters are there in the Bayesian network? What can you conclude about the expressive power of this type of network?

c. Now, consider a naive Bayes model where $X_1, \dots, X_n$ are evidence variables, and we have an additional class variable C, which has k possible values $c_1, \dots, c_k$. How many independent parameters are required to specify the naive Bayes model? How many independent parameters are required for an explicit representation of the joint distribution?

1). $P(X_{k+1} | X_1 \cdots X_k)$ need $l^k \cdot (l-1)$ params (k=3)



k variables that don't have parents

BN needs $(n-k)\, l^k (l-1)$ parameters in total

Joint distribution need $l^n - 1$ params

2). $\sum_{k=1}^{n-1} l^k (l-1) = (l-1) \cdot \frac{l(1-l^{n-1})}{1-l} = l^n - l$

almost same expressive power as full joint distribution

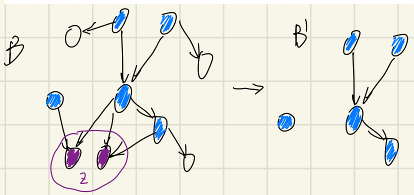
3). Naive Bayes need $(n-1) + k \cdot n$ params

joint distribution need $k \cdot 2^n$ params

3.8

Exercise 3.8

Let $\mathcal{B} = (\mathcal{G}, P)$ be a Bayesian network over some set of variables $\mathcal{X}$. Consider some subset of evidence nodes $\boldsymbol{Z}$, and let $\boldsymbol{X}$ be all of the ancestors of the nodes in $\boldsymbol{Z}$. Let $\mathcal{B}'$ be a network over the induced subgraph over $\boldsymbol{X}$, where the CPD for every node $X \in \boldsymbol{X}$ is the same in $\mathcal{B}'$ as in $\mathcal{B}$. Prove that the joint distribution over $\boldsymbol{X}$ is the same in $\mathcal{B}$ and in $\mathcal{B}'$. The nodes in $\mathcal{X} - \boldsymbol{X}$ are called *barren nodes* relative to $\boldsymbol{X}$, because (when not instantiated) they are irrelevant to computations concerning $\boldsymbol{X}$.



let $X_1 \cdots X_n$ be a topological ordering of X

$$P_{B'}(X_1 \cdots X_n) = \prod_{i=1}^{n} P(X_i | \text{parents}_{B'}(X_i))$$

$$P_B(X_1 \cdots X_n, Z) = \prod_{i=1}^{n} P(X_i | \text{parents}(X_i)) \cdot \prod_{z \in Z} P(z_i | \text{parents}(z_i))$$

$$P_B(X_1 \cdots X_n) = \frac{P_B(X_1 \cdots X_n, Z)}{P(Z | X_1 \cdots X_n)}$$

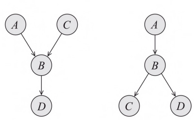
$$= \frac{\prod_{i=1}^{n} P(X_i | \text{parents}(X_i)) \cdot \prod_{z \in Z} P(z_i | \text{parents}(z_i))}{\prod_{z \in Z} P(z_i | \text{parents}(z_i))}$$

$$= P_{B'}(X_1 \cdots X_n)$$

3.15

Exercise 3.15

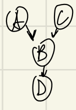
Consider the two networks:



(a) (b)

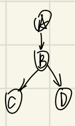
For each of them, determine whether there can be any other Bayesian network that is I-equivalent to it.

if G_1 and G_2 have the same skeleton and same v-structures, they are I-equivalent

1. 

$(A \perp C)$
$(A \perp D | B)$
$(C \perp D | B)$

no other.

2. 

$(C \perp D | B)$
$(A \perp C | B)$
$(A \perp D | B)$

$(C \perp D | B)$
$(A \perp C | B)$
$(A \perp D | B)$