52935 - Probabilistic Methods in Artificial Intelligence

Lecturer: Prof. Gal Elidan
Teaching Assistant: Ela Fallik

Hadar Tal

Hebrew University of Jerusalem, Israel

May 24, 2024

Contents

1 Probability Review **2**

2 Bayesian Networks **4**
- 2.1 Bayesian Networks Basics 4
- 2.2 Independencies and Factorization in Bayesian Networks 4
- 2.3 Reasoning Patterns in Bayesian Networks 5
- 2.4 D-separation and Global Markov Independencies 6
- 2.5 The relationship between Local and Global Markov Independencies 6
- 2.6 Markov Blanket in Bayesian Networks 7

3 Markov Chains **8**
- 3.1 Markov Chains Basics 8
- 3.2 Hidden Markov Models (HMMs) 8

4 Markov Networks **9**
- 4.1 Markov Networks Basics 9
- 4.2 Independencies and Factorization in Markov Networks 9
- 4.3 Markov Blanket in Markov Networks 10
- 4.4 Building a minimal I-map $\mathcal{H}$ for P 10
- 4.5 The transitions between Bayesian Networks and Markov Networks 11
  - 4.5.1 BM $\rightarrow$ MN 11
  - 4.5.2 MN $\rightarrow$ BM 11

1 Probability Review

Definition 1.1 (Probability Space)
A probability space is a triple $(\Omega, \mathcal{F}, P)$ where:

1. *Ω is the sample space*
2. *$\mathcal{F}$ is a σ-algebra of subsets of Ω*
3. *P is a probability measure on $\mathcal{F}$ such that $P(\Omega) = 1$*

Definition 1.2 (Joint Probability)
The joint probability of two events A and B is:

$$P(A, B) := P(A \cap B)$$

Definition 1.3 (Random Variable)
A random variable X is a function $X : \Omega \to \mathbb{R}$.
Val(X) = Image(X) = $\{x \in \mathbb{R} : \exists \omega \in \Omega \text{ s.t. } X(\omega) = x\}$

Definition 1.4 (Probability Mass Function (PMF))
The probability mass function of a random variable X is:

$$P(X = x) := P(\{\omega \in \Omega : X(\omega) = x\})$$

Definition 1.5 (Joint Distribution)
A joint distribution over a set of RVs $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ is a probability distribution $P_{\mathcal{X}} : Val(X_1) \times Val(X_2) \times \ldots \times Val(X_n) \to [0, 1]$ defined by:

$$\forall x_1, \ldots, x_n : x_i \in Val(X_i) \quad P_{\mathcal{X}}(x_1, x_2, \ldots, x_n) := P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Proposition 1.1 (Law of Total Probability)
For X, Y random variables, we can write:

$$P(X) = \sum_{y \in Val(Y)} P(X, Y = y)$$

Definition 1.6 (Conditional distribution)
For X, Y RVs, and for any $y \in Val(Y)$ where $P(Y = y) > 0$ the conditional distribution of X given Y=y is:

$$P(X|y) := \frac{P_{X,Y}(X = x, Y = y)}{P_Y(Y = y)}$$

Proposition 1.2 (Chain Rule)
For any set of random variables $X_1, X_2, \ldots, X_n$:

$$P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \ldots P(X_n|X_1, X_2, \ldots, X_{n-1})$$

Proposition 1.3 (Bayes' Rule)
For any two random variables H, E:

$$P(H = h|E = e) = \frac{P(E = e|H = h)P(H = h)}{P(E = e)}$$

where we often call:

- *$P(H = h)$ the **prior** probability*

- $P(H = h|E = e)$ *the* ***posterior*** *probability in light of evidence* $E = e$
- $P(E = e|H = h)$ *the* ***likelihood*** *of the evidence* $E = e$ *given the hypothesis* $H = h$

Definition 1.7 (Marginal Independence)
Let P *be a probability distribution over a set of random variables* $\mathcal{X}$ *and let* $X, Y \in \mathcal{X}$. *We say that* X *is independent of* Y, *denoted* $P \models X \perp Y$, *if*

$$P(X|Y) = P(X)$$

Definition 1.8 (Conditional Independence)
Let P *be a probability distribution over a set of random variables* $\mathcal{X}$ *and let* $X, Y, Z \in \mathcal{X}$. *We say that* X *is independent of* Y *given* Z, *denoted* $P \models X \perp Y|Z$, *if*

$$P(X|Y, Z) = P(X|Z)$$

Lemma 1.1 (Equivalent Definitions of Conditional Independence)
Let P *be a probability distribution over a set of random variables* $\mathcal{X}$ *and let* $X, Y, Z \in \mathcal{X}$. *The following are equivalent:*

1. $P \models X \perp Y|Z$
2. $P(X, Y|Z) = P(X|Z)P(Y|Z)$
3. $P(X, Y, Z) = P(X|Z)P(Y, Z)$
4. $\exists f, g : P(X, Y, Z) = f(X, Z)g(Y, Z)$

Theorem 1.1 (Properties of Conditional Independence)
Let P *be a probability distribution over a set of random variables* $\mathcal{X}$ *and let* $X, Y, Z, W \in \mathcal{X}$. *The following hold:*

1. ***Symmetry*** - $(X \perp Y|Z) \implies (Y \perp X|Z)$
2. ***Decomposition*** - $(X \perp Y, W|Z) \implies (X \perp Y|Z) \wedge (X \perp W|Z)$
3. ***Weak Union*** - $(X \perp Y, W|Z) \implies (X \perp Y|W, Z)$
4. ***Contraction*** - $(X \perp Y|Z) \wedge (X \perp W|Y, Z) \implies (X \perp Y, W|Z)$
5. ***Intersection*** - *For strictly positive distributions,*

$$(X \perp Y|W, Z) \wedge (X \perp W|Y, Z) \implies (X \perp Y, W|Z)$$

2 Bayesian Networks

2.1 Bayesian Networks Basics

Definition 2.1 (Probabilistic Graphical Model (PGM))
A probabilistic graphical model is a pair $(\mathcal{G}, P)$ where:

1. *$\mathcal{G}$ is a graph*
2. *P is a probability distribution*

Definition 2.2 (Bayesian Network)
A Bayesian Network $\mathcal{B}$ is:

1. ***Bayesian Network Structure** - A directed acyclic graph (DAG) $\mathcal{G} = (\mathcal{X}, E)$ $(|\mathcal{X}| = n)$*
2. ***Set of CPDs** - $\{P_i(X_i|Pa(X_i))\}_{i=1}^{n}$*

the network defines a probability distribution:

$$P_{\mathcal{B}}(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P_i(X_i|Pa(X_i))$$

A Bayesian Network is the tuple $\mathcal{B} = (\mathcal{G}, P_{\mathcal{B}})$.

Theorem 2.1 (Bayesian Network defines a probability distribution)
For any Bayesian Network B, $P_B(X_1, X_2, \ldots, X_n)$ is a joint probability distribution over the variables $X_1, X_2, \ldots, X_n$.

Definition 2.3 (Descendants of a node)
Let $G = (V, E)$ be a directed graph and let $X_i \in V$. The descendants of X_i are:

$$D(X_i) = \{X_j \in \mathcal{X} : \exists \text{ directed path } X_i \to \cdots \to X_j\}$$

Definition 2.4 (Naive Bayes Model)
A Naive Bayes Model is a Bayesian Network where all the features are non adjacent children of the class node.

Definition 2.5 (Naive Bayes Classifier)
A Naive Bayes Classifier is a classifier that uses the Naive Bayes Model to classify instances.

$$\hat{c} = \operatorname*{argmax}_{c \in C} P(c|x_1, x_2, \ldots, x_n) = \operatorname*{argmax}_{c \in C} P(c, x_1, x_2, \ldots, x_n) = \operatorname*{argmax}_{c \in C} P(c) \prod_{i=1}^{n} P(x_i|c)$$

2.2 Independencies and Factorization in Bayesian Networks

Definition 2.6 ($I_{LM}(\mathcal{G})$)
*The **Local Markov Independencies Set** of a Bayesian Network B is the set of all independencies that hold in the network:*

$$I_{LM}(\mathcal{G}) = \{(X_i \perp ND(X_i)|Pa(X_i))\}_{i=1}^{|\mathcal{X}|}$$

Definition 2.7 (I(P))
The set of independencies that hold in a distribution P over $\mathcal{X}$ is:

$$I(P) = \{(X \perp Y|Z) : (X, Y, Z) \subseteq \mathcal{X}, \quad P \models (X \perp Y|Z)\}$$

Definition 2.8 (I-map)
A DAG $\mathcal{G}$ is an I-map of a distribution P if all independencies assumptions of $\mathcal{G}$ hold in P:

$$I_{LM}(\mathcal{G}) \subseteq I(P)$$

Theorem 2.2 (Factorization)
If G is an I-map of P, then we can write:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa(X_i))$$

Definition 2.9 (Factorization)
We say that P factorizes over $\mathcal{G}$ if there exist CPDs $\{P_i\}_{i=1}^{n}$ such that:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P_i(X_i | Pa(X_i))$$

Corollary 2.1 (Independencies implies Factorization)
If $\mathcal{G}$ is an I-map of P ($P \models I_{LM}(\mathcal{G})$), then P factorizes over $\mathcal{G}$.

Corollary 2.2 (Independencies implies Factorization (2))
If $\mathcal{G}$ is an I-map of P ($P \models I_{LM}(\mathcal{G})$), then $(\mathcal{G}, P)$ is a Bayesian Network.

Theorem 2.3 (Independencies in P_B)
For $P_{\mathcal{B}}$ it holds for all i that

1. $X_i \perp ND(X_i) | Pa(X_i) \quad (I_{LM}(\mathcal{G}))$
2. $P_{\mathcal{B}}(X_i | ND(X_i)) = P_i(X_i | Pa(X_i))$

Corollary 2.3 (Factorization implies Independencies)
If P factorizes over $\mathcal{G}$, then $\mathcal{G}$ is an I-map of P ($P \models I_{LM}(\mathcal{G})$).

Theorem 2.4 (Fundmental Theorem of Bayesian Networks)
Let $\mathcal{G}$ be a BN structure over $\mathcal{X} = X_1, X_2, \ldots, X_n$ and let P be a joint distribution over $\mathcal{X}$. Then $\mathcal{G}$ is an I-map of $P \Leftrightarrow P$ factorizes over $\mathcal{G}$.

Definition 2.10 (Minimal I-map)
A DAG $\mathcal{G}$ is a minimal I-map of a distribution P if

1. $\mathcal{G}$ *is an I-map of* P
2. *If* $\mathcal{G}' \subset \mathcal{G}$ *then* $\mathcal{G}'$ *is not an I-map of* P

2.3 Reasoning Patterns in Bayesian Networks

Definition 2.11 (Reasoning Patterns in Bayesian Networks)
There are 4 main reasoning patterns in Bayesian Networks:

- ***Downstream (causal) reasoning*** - $X \to Z \to Y$
- ***Upstream (evidential) reasoning*** - $X \leftarrow Z \leftarrow Y$
- ***Common Causal reasoning*** - $X \leftarrow Z \to Y$
- ***Common Effect reasoning*** - $X \to Z \leftarrow Y$

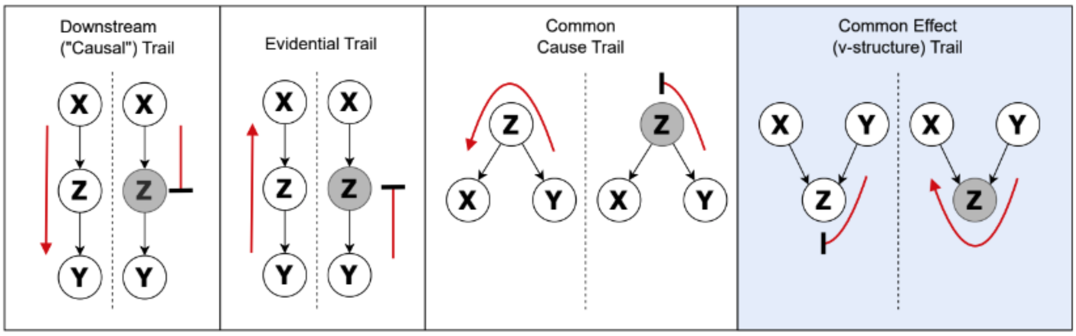

Figure 1: Reasoning Patterns in Bayesian Networks

2.4 D-separation and Global Markov Independencies

Question 2.1
If P factorizes over $\mathcal{G}$, then $\mathcal{G}$ is an I-map of P ($P \models I_{LM}(\mathcal{G})$).
Can p satisfy more independencies than those implied by $\mathcal{G}$? Yes.
Given $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathcal{X}$, we would like to characterize when does $P \models I_{LM}(\mathcal{G}) \implies P \models X \perp Y|Z$.
Or characterize the complement - Can we find P that factorizes over $\mathcal{G}$ but $P \not\models X \perp Y|Z$?

Definition 2.12 (Active Trail)
A trail $X = X_1 - X_2 - \cdots - X_n$ between X and Y in a BN is active given a set of observed RVs Z, if whenever there is a v-structure along the trail $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, then X_i or one of its descendants is in Z, and all other nodes along the trail are not in Z.

Definition 2.13 (d-separation)
The sets $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated given $\boldsymbol{Z}$ in $\mathcal{G}$, denoted d-sep$_{\mathcal{G}}(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{Z})$, if there is no active trail between any node in $\boldsymbol{X}$ and any node in $\boldsymbol{Y}$ given $\boldsymbol{Z}$.

Definition 2.14 (Global Markov Independencies)
The set of global Markov Independencies of a BN structure $\mathcal{G}$ is the set of all independencies that correspond to d-separation:

$$I(\mathcal{G}) := I_{GM}(\mathcal{G}) := \{(X \perp Y|Z) : d\text{-}sep_{\mathcal{G}}(X;Y|Z)\}$$

d-separation characterizes precisely the full set of independencies that a BN structure encodes.

Theorem 2.5 (Soundness)
If a distribution P factorizes over a BN structure $\mathcal{G}$, then $I(\mathcal{G}) \subseteq I(P)$.

Note - the other direction is not true. If a distribution P factorizes over $\mathcal{G}$, then it is not necessarily true that $I(P) \subseteq I(\mathcal{G})$.

Theorem 2.6 (Completness)
If $(X \perp Y|Z) \notin I(\mathcal{G})$, then there exists a distribution P that factorizes over $\mathcal{G}$ in which $P \not\models (X \perp Y|Z)$.

2.5 The relationship between Local and Global Markov Independencies

Proposition 2.1 ($I_{LM}(\mathcal{G}) \subseteq I_{GM}(\mathcal{G})$)
For any BN structure $\mathcal{G}$, it holds that $I_{LM}(\mathcal{G}) \subseteq I_{GM}(\mathcal{G})$.

Proposition 2.2
For a DAG $\mathcal{G}$ and distribution P, it holds that $P \models I_{LM}(\mathcal{G}) \iff P \models I_{GM}(\mathcal{G})$.

Definition 2.15 (Perfect Map (P-map))
A graph $\mathcal{G}$ is a P-map of a distribution P if $I(\mathcal{G}) = I(P)$.

2.6 Markov Blanket in Bayesian Networks

Definition 2.16 (Markov Blanket in a Bayesian Network Structure)
The Markov Blanket of a node X_i in a BN structure $\mathcal{G}$ is the mininal set of nodes that renders X_i independent of all other nodes in the graph

$$MB(X_i) := \underset{\boldsymbol{Z} \in \mathcal{X}}{\operatorname{argmin}} \{|\boldsymbol{Z}| : X_i \perp \mathcal{X}/(X_i \cup \boldsymbol{Z}) | \boldsymbol{Z}\}$$

Definition 2.17 (Spouses of a node)
The spouses of a node X_i in a BN structure $\mathcal{G}$ are the parents of the children of X_i

$$Sp(X_i) := \{X_j \in \mathcal{X} : X_j \textit{ is a parent of a child of } X_i\}$$

Theorem 2.7 (Markov Blanket Theorem (in Bayesian Networks))
For any node X_i in a BN structure $\mathcal{G}$, it holds that:

$$MB(X_i) = Pa(X_i) \cup Ch(X_i) \cup Sp(X_i)$$

Lemma 2.1 (Markov Blanket is not unique)
Let P be a distribution over $\mathcal{X}$ and let $X_i \in \mathcal{X}$. $MB(X_i)$ is not unique.

3 Markov Chains

3.1 Markov Chains Basics

Definition 3.1 (Markov Property)
Given a random process $\{X^{(t)}\}_{t\in\mathbb{N}}$, over a state space $\mathcal{S}$, we say it possess the Markov property, if $\forall t \geq 1$,

$$X^{(t+1)} \perp X^{(1)}, \ldots, X^{(t-1)} | X^{(t)}$$

that is,

$$\forall t \in \mathbb{N} \quad P_t(X^{(t+1)}|X^{(1)}, \ldots, X^{(t)}) = P(X^{(t+1)}|X^{(t)})$$

Definition 3.2 (Time-invariant assumption)

$$\forall t \in \mathbb{N} \quad P_t(X^{(t+1)}|X^{(t)}) = P(X^{(t+1)}|X^{(t)})$$

Definition 3.3 (Markov Chain)
A Markov Chain is a random process $\{X^{(t)}\}_{t\in\mathbb{N}}$ that possess the Markov property.

Note: Markov Chains are often represented as a directed graph, where each node is a state in the state space, and each edge is a transition between states. The transition diagram is a very different entity from the graph encoding the structure of a Bayesian Network.

3.2 Hidden Markov Models (HMMs)

Definition 3.4 (Hidden Markov Model)
A pair of random processes $(\{X^{(t)}\}_{t\in\mathbb{N}}$ and $\{O^{(t)}\}_{t\in\mathbb{N}})$ is a Hidden Markov Model if:

1. *$X^{(t)}$ is a Markov process whose behavior is not directly observable (hidden).*
2. *For each t,*

$$O^{(t)} \perp X^{(1:t-1)}, X^{(t+1:T)} | X^{(t)}$$

   that is,

$$P(O^{(t)}|X^{(1:t-1)}, X^{(t+1:T)}) = P(O^{(t)}|X^{(t)})$$

We call:

- *$X^{(t)}$ the **hidden states**,*
- *$O^{(t)}$ the **observed states**,*
- *$P(X^{(t+1)}|X^{(t)})$ the **transition probability***
- *$P(O^{(t)}|X^{(t)})$ the **emission probability***

4 Markov Networks

4.1 Markov Networks Basics

Definition 4.1 (Gibbs Distribution)
A Gibbs distribution over a set of random variables $\mathcal{X}$ is a probability distribution of the form:

$$P_{\mathcal{X}}(x_1, x_2, \ldots, x_n) = \frac{1}{\mathcal{Z}} \prod_j \phi_j(X_{c_j})$$

where

- $\mathcal{C} = \{c_1, c_2, \ldots, c_m\}$ *is a set of cliques in the graph.*
- X_{c_j} *is the set of random variables in clique* c_j
- $\phi_j : Val(X_{c_j}) \to \mathbb{R}^+$ *is a potential function / factor over the clique* c_j
- $\mathcal{Z}$ *is the normalization constant (partition function)*

Definition 4.2 (Markov Network)
A Markov Network $\mathcal{M}$ *is:*

1. ***Markov Network Structure*** *- An undirected graph* $\mathcal{H} = (\mathcal{X}, E)$ $(|\mathcal{X}| = n)$
2. ***Set of Factors*** *-* $\Phi = \{\phi_j(X_{c_j})\}_j$ *such that every* X_{c_j} *is a clique in* $\mathcal{H}$

the network defines a probability distribution:

$$P_{\Phi}(X_1, X_2, \ldots, X_n) = \frac{1}{\mathcal{Z}} \prod_j \phi_j(X_{c_j})$$

A Markov Network is the tuple $\mathcal{M} = (\mathcal{H}, P_{\Phi})$.

4.2 Independencies and Factorization in Markov Networks

Definition 4.3 (Seperating Set)
A set of nodes Z *separates* X *and* Y *in an undirected graph* $\mathcal{H}$ *if every path between* X *and* Y *passes through* Z*, denoted* $sep_{\mathcal{H}}(X; Y|Z)$.

Definition 4.4 (I($\mathcal{H}$))
The set of independencies that hold in a MN structure $\mathcal{H}$ *is*

$$I(\mathcal{H}) = \{(X \perp Y|Z) : sep_{\mathcal{H}}(X; Y|Z)$$

Lemma 4.1 (Monotonicity of seperation)
If $Z \subseteq Z'$*, then* $sep_{\mathcal{H}}(X; Y|Z) \subseteq sep_{\mathcal{H}}(X; Y|Z')$.

Definition 4.5 (Factorization)
We say that P *factorizes over* $\mathcal{H}$ *if there exist factors* $\{\phi_j\}_j$ *and a normalization constant* $\mathcal{Z}$ *such that:*

$$P(X_1, X_2, \ldots, X_n) = \frac{1}{\mathcal{Z}} \prod_j \phi_j(X_{c_j})$$

such that every c_j *is a clique in* $\mathcal{H}$.

Theorem 4.1 (Soundness)
If a distribution P *factorizes over* $\mathcal{H}$*, then* $\mathcal{H}$ *is an I-map of* P $(P \models I(\mathcal{H}))$.

Theorem 4.2 (Completeness)
Let $\mathcal{H}$ be a MN structure. If X, Y are not separated by Z in $\mathcal{H}$, then there exists a distribution P that factorizes over $\mathcal{H}$ in which $P \not\models X \perp Y|Z$.

Definition 4.6 ($I_{pair}(\mathcal{H})$)
*The **Pairwise Independencies Set** of a $\mathcal{H}$ is defined as:*

$$I_{pair}(\mathcal{H}) = \{(X \perp Y|\mathcal{X}/\{X, Y\}) : X - Y \notin \mathcal{H}\}$$

(i.e., X and Y are independent given all other nodes)

Definition 4.7 (Neighbors of a node ($Ne(X_i)$))
The neighbors of a node X_i in a MN structure $\mathcal{H}$ are the nodes that are connected to X_i in the graph

$$Ne(X_i) := \{X_j \in \mathcal{X} : X_i - X_j \in \mathcal{H}\}$$

Definition 4.8 ($I_{local}(\mathcal{H})$)
*The **Local Independencies Set** of a $\mathcal{H}$ is defined as:*

$$I_{local}(\mathcal{H}) = \{(X_i \perp (\mathcal{X}/(\{X_i\} \cup Ne(X_i)))|Ne(X_i))\}$$

(i.e., X_i is independent of all other nodes given its neighbors)

Lemma 4.2 (Independence implies pairwise independence)

$$P \models I(\mathcal{H}) \implies P \models I_{local}(\mathcal{H}) \implies I_{pair}(\mathcal{H})$$

Theorem 4.3 (For positive P, pairwise independence implies independence)
If P is a strictly positive distribution, then

$$P \models I_{pair}(\mathcal{H}) \Leftrightarrow P \models I(\mathcal{H})$$

4.3 Markov Blanket in Markov Networks

Definition 4.9 (Markov Blanket in a Markov Network Structure)
The Markov Blanket of a node X_i in a MN structure $\mathcal{H}$ is the mininal set of nodes that renders X_i independent of all other nodes in the graph

$$MB(X_i) := \underset{\boldsymbol{Z} \in \mathcal{X}}{\text{argmin}}\{|\boldsymbol{Z}| : X_i \perp \mathcal{X}/(X_i \cup \boldsymbol{Z})|\boldsymbol{Z}\} = \underset{\boldsymbol{Z} \in \mathcal{X}}{\text{argmin}}\{|\boldsymbol{Z}| : sep_{\mathcal{H}}(X_i; \mathcal{X}/(X_i \cup \boldsymbol{Z})|\boldsymbol{Z})\}$$

Theorem 4.4 (Markov Blanket Theorem (in Markov Networks))
For strictly positive MN $(\mathcal{H}, P_\Phi)$, for any node X_i in $\mathcal{H}$, it holds that:

$$MB(X_i) = Ne(X_i)$$

4.4 Building a minimal I-map $\mathcal{H}$ for P

Definition 4.10 (I-map)
An undirected graph $\mathcal{H}$ is an I-map of a distribution P if all independencies assumptions of $\mathcal{H}$ hold in P:

$$I(\mathcal{H}) \subseteq I(P)$$

Algorithm 1: Building a minimal I-map $\mathcal{H}$ for P

Input: An oracle that returns $\forall X, Y, Z \subseteq \mathcal{X}$ if $P \models X \perp Y | Z$

Output: $\mathcal{H}$ - a minimal I-map of P

```
foreach X_i ∈ X_1, ..., X_n do
    foreach X_j ∈ {X_1, ..., X_n}/{X_i} do
        if P ⊭ X_i ⊥ X_j | 𝒳/{X_i, X_j} then
            H_e ← H_e ∪ {X_i − X_j}
        end
    end
end
```

Theorem 4.5 (Uniqueness of minimal I-map)
There exists a unique undirected minimal I-map $\mathcal{H}$ of a distribution P.

4.5 The transitions between Bayesian Networks and Markov Networks

4.5.1 BM $\rightarrow$ MN

Definition 4.11 (Immorality)
An immorality in a BN structure $\mathcal{G}$ is a pair of nodes X_i, X_j that have a common child X_k but are not connected by an edge.

Definition 4.12 (The Moral Graph M($\mathcal{G}$))
The moral graph of a BN structure $\mathcal{G}$ is the undirected graph that is created in the following way:

1. *$M(\mathcal{G})$ contains all the nodes of $\mathcal{G}$*
2. *$M(\mathcal{G})$ contains all the edges in $\mathcal{G}$ undirected*
3. *For every pair of nodes X_i, X_j that have a common child in $\mathcal{G}$, add an edge between X_i and X_j in $M(\mathcal{G})$*

Theorem 4.6 ($M(\mathcal{G})$ is a minimal I-map of G)
The moral graph $M(\mathcal{G})$ is a minimal I-map of the distribution P that has exactly the same independencies as $\mathcal{G}$.

Theorem 4.7
If $\mathcal{G}$ is a moral graph, then $M[\mathcal{G}]$ is a P-map of G; i.e.,

$$I(\mathcal{G}) = I(M[\mathcal{G}])$$

4.5.2 MN $\rightarrow$ BM

Definition 4.13 (Triangulated Graph)
A graph $\mathcal{H}$ is triangulated if every cycle of length ≥ 4 has a chord.

Definition 4.14 (Skeleton of a Bayesian Network)
The skelleton of a Bayesian Network $\mathcal{G}$ is the undirected graph that is created by removing the directions of the edges in $\mathcal{G}$.

Theorem 4.8
If $\mathcal{G}$ is an I-map of $\mathcal{H}$, then the skeleton of $\mathcal{G}$ is a triangulation of $\mathcal{H}$.

Lemma 4.3
If $\mathcal{G}$ is a moral graph, then the skeleton of $\mathcal{G}$ is a triangulated graph.