Markov Networks

Or undirected graphical models

Wilker Aziz

w.aziz@uva.nl

Fall 2025 (v2)

https://probabll.github.io

Module 2 introduces *Markov networks* (MNs; Chapter 4).

ILOs After this module the student

- can map MNs to distributions and vice-versa;
- can reason using an MN;
- recognises the flow of probabilistic influence in MNs;
- recognises independence and separation.

Textbook for this course: Koller and Friedman [1].

HC2b: MNs – semantics.

HC3a: MNs – reasoning.

LC3: MNs in code.

HC3b: MNs – influence.

WC3: exercises (semantics, reasoning and influence).

Table of contents

Semantics

We will introduce *Markov Networks* (MNs; Chapter 4) via an example from the textbook (the *Misconception* example from Section 4.1).

After presenting the complete example, we will introduce MNs in full generality.

Four students (Alice, Bob, Charles and Debbie) work together in pairs.

Alice and Charles are not on speaking terms, nor are Bob and Debbie. Only the following pairs study together:

- Alice and Bob
- Bob and Charles
- Charles and Debbie
- Debbie and Alice

In class, the professor accidentally misspoke, giving rise to a misconception. After class, the students may figure out the problem and share their newfound understanding with their study partners.

The rv A denotes whether Alice has the misconception a^1 or not a^0. We use B, C, D analogously for the other students.

Because Alice and Charles never speak, we have that *A* and *C* are conditionally independent given *B* and *D*. $A \perp C \mid B, D$

Similarly, *B* and *D* are conditionally independent given *A* and *C*.

$B \perp D \mid A, C$

Can we represent these two statements with a BN?

Because Alice and Charles never speak, we have that *A* and *C* are conditionally independent given *B* and *D*. $A \perp C \mid B, D$

Similarly, *B* and *D* are conditionally independent given *A* and *C*.

$B \perp D \mid A, C$

Can we represent these two statements with a BN?

Attempt 1:

- $A \perp C \mid B, D$ ✓
- but D, B are independent given A alone ✗
- and if we are given A, C then D, B become dependent ✗
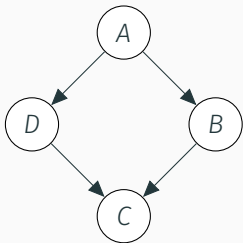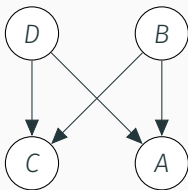
Because Alice and Charles never speak, we have that *A* and *C* are conditionally independent given *B* and *D*. $A \perp C \mid B, D$

Similarly, *B* and *D* are conditionally independent given *A* and *C*.

$$B \perp D \mid A, C$$

Can we represent these two statements with a BN?

Attempt 2:

- $A \perp C \mid B, D$ ✓
- but now *B* and *D* are marginally independent ✗

Undirected Interaction

We want to represent two conditional independences $A \perp C \mid B, D$
and $B \perp D \mid A, C$ **and no other**.

It turns out that no BN structure can do precisely that. Instead,

- we need to represent the interaction between pairs
- and we have no reason to ascribe direction to any interaction
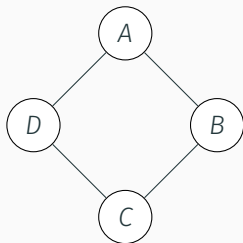
We need an **undirected**
representation of the relevant
interactions!

Figure 1: The MN graph for the
Misconception example

Undirected Interaction

We want to represent two conditional independences $A \perp C \mid B, D$ and $B \perp D \mid A, C$ **and no other**.

It turns out that no BN structure can do precisely that. Instead,

- we need to represent the interaction between pairs
- and we have no reason to ascribe direction to any interaction

Figure 1: The MN graph for the *Misconception* example
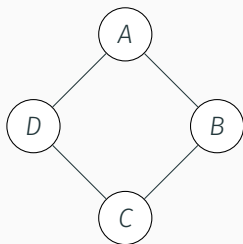
We need an **undirected** representation of the relevant interactions! But how can we parameterise a joint distribution coherent with this view?

A factor ϕ with scope *X*, *Y* represents the undirected *affinity* between possible outcomes of *X* and *Y*.

A factor ϕ with scope X, Y represents the undirected *affinity* between possible outcomes of X and Y.

	$\phi_1(A, B)$	
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

A factor ϕ with scope X, Y represents the undirected *affinity* between possible outcomes of X and Y.

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

A factor ϕ with scope X, Y represents the undirected *affinity* between possible outcomes of X and Y.

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_3(C, D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

A factor ϕ with scope X, Y represents the undirected *affinity* between possible outcomes of X and Y.

$\phi_4(D, A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_3(C, D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

A factor ϕ with scope X, Y represents the undirected *affinity* between possible outcomes of X and Y.

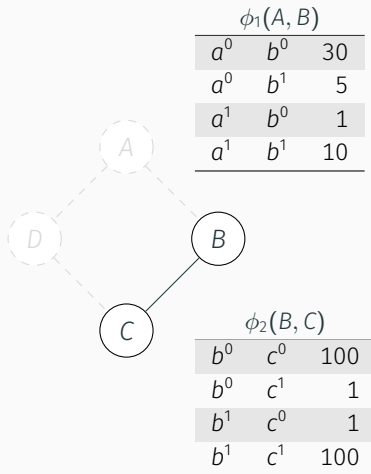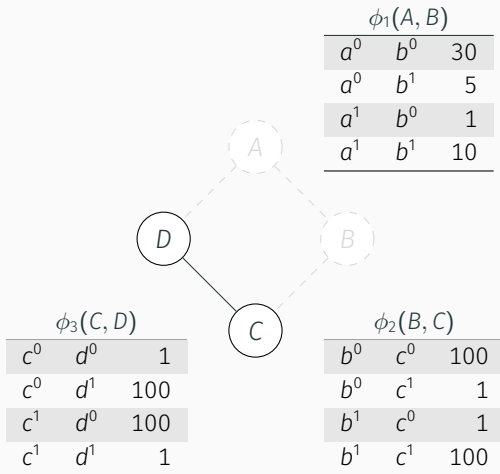| $\phi_4(D, A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_3(C, D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

Figure 2: Factors for the *Misconception* example

A factor ϕ is a function from the joint outcome space of a set D of rvs, which we call the factor's *scope*, to $\mathbb{R}_{\geq 0}$. Factors represent affinity amongst interacting variables.

Think of a factor ϕ as something like a python function with named arguments, the order of the arguments doesn't matter because they are 'named' with the names of the rvs in Scope[ϕ].

A factor ϕ is a function from the joint outcome space of a set D of rvs, which we call the factor's *scope*, to $\mathbb{R}_{\geq 0}$. Factors represent affinity amongst interacting variables.

$\phi_4(D, A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_3(C, D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100
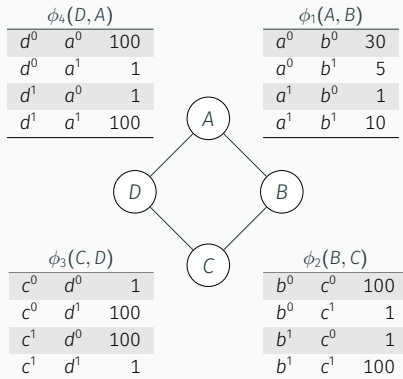
- Bob is more strongly inclined to agree with Charles than with Alice.
- Charles and Debbie on the other hand are strongly inclined to disagreement.

Think of a factor ϕ as something like a python function with named arguments, the order of the arguments doesn't matter because they are 'named' with the names of the rvs in Scope[ϕ].
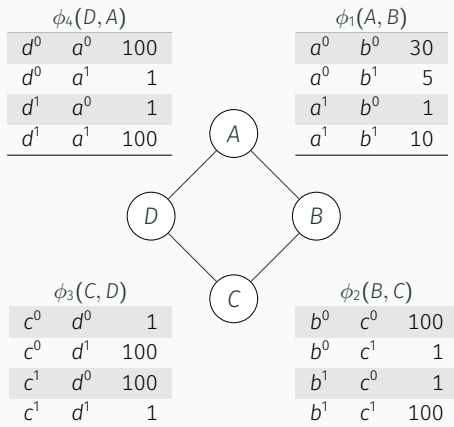
9

$\phi_4(D, A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_3(C, D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

1. $P(A = a^0, B = b^0, C = c^0, D = d^0) =$
2. $P(A = a^1, B = b^0, C = c^0, D = d^1) =$

A *joint probability distribution* is a means to represent our uncertainty over the *possible* assignments of the rvs of interest.

$\phi_4(D, A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10
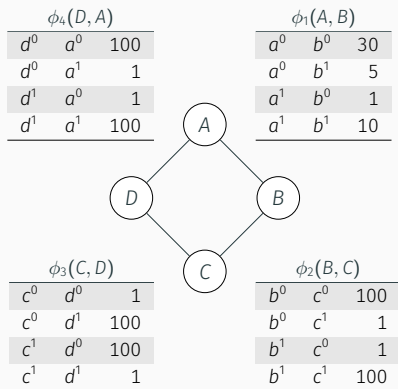
With its undirected graph and factors, an MN identifies one such distribution.

$\phi_3(C, D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

The probability of any one of the possible outcomes of (A, B, C, D) in *Misconception* is proportional to $\phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A)$.

In an MN, we define a distribution by taking the **product of the local factors** and then **normalising it** to define a valid joint distribution:
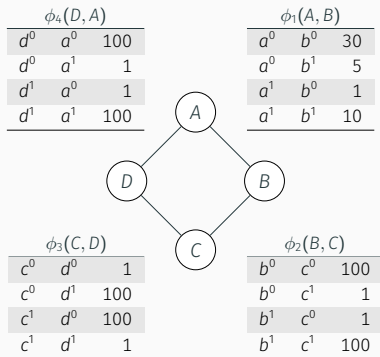
A	B	C	D	Unnormalised product of factors	$P(A, B, C, D)$
a^0	b^0	c^0	d^0	300 000	0.04
a^0	b^0	c^0	d^1	300 000	0.04
a^0	b^0	c^1	d^0	300 000	0.04
a^0	b^0	c^1	d^1	30	4.1×10^{-6}
a^0	b^1	c^0	d^0	500	6.9×10^{-5}
a^0	b^1	c^0	d^1	500	6.9×10^{-5}
a^0	b^1	c^1	d^0	5 000 000	0.69
a^0	b^1	c^1	d^1	500	6.9×10^{-5}
a^1	b^0	c^0	d^0	100	1.4×10^{-5}
a^1	b^0	c^0	d^1	1 000 000	0.14
a^1	b^0	c^1	d^0	100	1.4×10^{-5}
a^1	b^0	c^1	d^1	100	1.4×10^{-5}
a^1	b^1	c^0	d^0	10	1.4×10^{-6}
a^1	b^1	c^0	d^1	100 000	0.014
a^1	b^1	c^1	d^0	100 000	0.014
a^1	b^1	c^1	d^1	100 000	0.014
Sum				7 201 840	1.0

Table 1: Joint Distribution over $\mathsf{Val}(A) \times \mathsf{Val}(B) \times \mathsf{Val}(C) \times \mathsf{Val}(D)$

The BN indeed *identifies* the joint distribution, but not by independently representing one probability value per joint outcome.

The probability value of any one of the joint outcomes is expressed via a product of local factors for directly interacting variables.
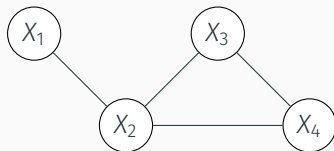
$\phi_4(D, A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_3(C, D)$		
c^0	d^1	1
c^0	d^0	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

If each of N rvs takes on one of K values, we infer all K^N joint probabilities from the affinities of N K-by-K-dimensional factors.

While in this example (K=2, N=4), K^N happens to be equal to NK^2, for larger K and N the MN is very compact.

13

The notion of *directly interacting variables* corresponds to what is known as **complete subgraphs**. The factors in the MN, by definition, have complete subgraphs as their scope.
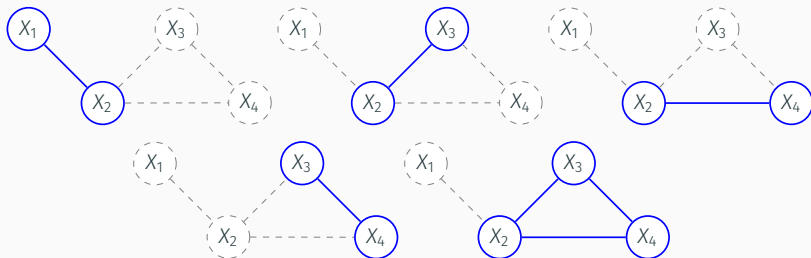

Some PGM jargon: we say that a factor 'covers' the rvs in its scope, and we also say that a factor 'covers' the edges that connect any two rvs in its scope.
Definition 2.13. In a complete subgraph, also known as *clique*, every two nodes are connected by some edge.
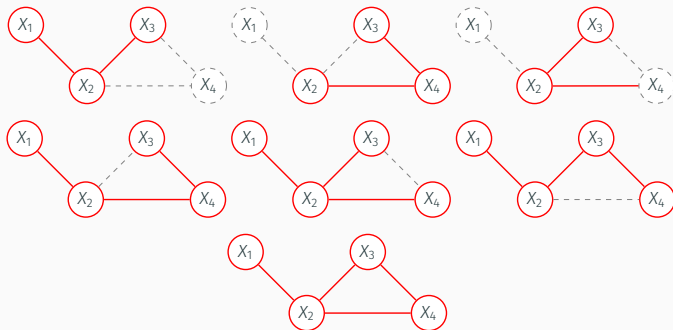
The notion of *directly interacting variables* corresponds to what is known as **complete subgraphs**. The factors in the MN, by definition, have complete subgraphs as their scope.

Some PGM jargon: we say that a factor 'covers' the rvs in its scope, and we also say that a factor 'covers' the edges that connect any two rvs in its scope.
Definition 2.13. In a complete subgraph, also known as *clique*, every two nodes are connected by some edge.

The notion of *directly interacting variables* corresponds to what is known as **complete subgraphs**. The factors in the MN, by definition, have complete subgraphs as their scope.

Some PGM jargon: we say that a factor 'covers' the rvs in its scope, and we also say that a factor 'covers' the edges that connect any two rvs in its scope.

Definition 2.13. In a complete subgraph, also known as *clique*, every two nodes are connected by some edge.

14

An MN is a graph $\mathcal{H}$ and a collection Φ of factors such that:

- $\mathcal{H}$ is an undirected graph;
- nodes in $\mathcal{H}$ represent the rvs $X_1, \ldots, X_n$;
- the edges in $\mathcal{H}$ indicate direct interaction between rvs;
- the scope $\mathbf{D}_i$ of a factor $\phi_i \in \Phi$ is a complete subgraph of $\mathcal{H}$;
- Φ covers all nodes and all edges in the graph.

Z is also known as the *partition function* (a term that comes from statistical physics).

An MN is a graph $\mathcal{H}$ and a collection Φ of factors such that:

- $\mathcal{H}$ is an undirected graph;
- nodes in $\mathcal{H}$ represent the rvs $X_1, \ldots, X_n$;
- the edges in $\mathcal{H}$ indicate direct interaction between rvs;
- the scope $\mathbf{D}_i$ of a factor $\phi_i \in \Phi$ is a complete subgraph of $\mathcal{H}$;
- Φ covers all nodes and all edges in the graph.

The MN represents a joint distribution by normalisation of the product of factors Φ:

$$P_\Phi(X_1, \ldots, X_n) = \frac{1}{Z} \prod_{\phi_i \in \Phi} \phi_i(\mathbf{D}_i) \tag{1}$$

where $\tilde{P}_\Phi(X_1, \ldots, X_n) = \prod_{\phi_i \in \Phi} \phi_i(\mathbf{D}_i)$ is an unnormalised measure, and $Z = \sum_{X_1, \ldots, X_n} \tilde{P}_\Phi(X_1, \ldots, X_n)$ is the constant that normalises it.
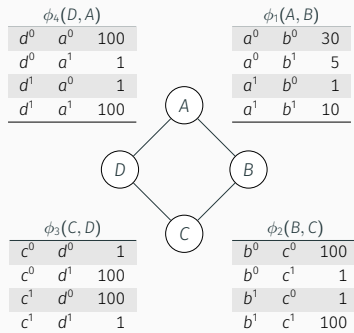

Z is also known as the *partition function* (a term that comes from statistical physics).

15

With its undirected graph $\mathcal{H}$ and a collection of factors Φ, an MN identifies a factorised distribution P_Φ.

In the *Misconception* example:

$\phi_4(D, A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_3(C, D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

Φ	$\mathbf{D}_i = \text{Scope}[\phi_i]$
ϕ_1	$\{A, B\}$
ϕ_2	$\{B, C\}$
ϕ_3	$\{C, D\}$
ϕ_4	$\{D, A\}$

Factors and their scope

$$P_\Phi(A, B, C, D) = \frac{1}{Z} \prod_{\phi_i \in \Phi} \phi_i(\mathbf{D}_i) = \frac{1}{Z}\phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

This is similar for BNs, whose DAG and CPDs identify a factorised distribution. 16

Since we defined factors as functions whose scopes are complete subgraphs (any two rvs in the scope must be connected by an edge), a collection of factors Φ identifies a graph $\mathcal{H}$ uniquely.

Can you identify the graph in each case?

Misconception	
Φ	$\mathbf{D}_i = \mathrm{Scope}[\phi_i]$
ϕ_1	$\{A, B\}$
ϕ_2	$\{B, C\}$
ϕ_3	$\{C, D\}$
ϕ_4	$\{D, A\}$

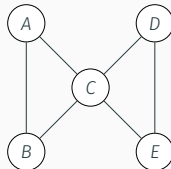| Common Friend	
Φ	$\mathbf{D}_i = \mathrm{Scope}[\phi_i]$
ϕ_1	$\{A, B\}$
ϕ_2	$\{A, C\}$
ϕ_3	$\{B, C\}$
ϕ_4	$\{C, D, E\}$

Since we defined factors as functions whose scopes are complete subgraphs (any two rvs in the scope must be connected by an edge), a collection of factors Φ identifies a graph $\mathcal{H}$ uniquely.

Can you identify the graph in each case?

	Misconception			Common Friend	
Φ	$\mathbf{D}_i = \text{Scope}[\phi_i]$		Φ	$\mathbf{D}_i = \text{Scope}[\phi_i]$	
ϕ_1	$\{A, B\}$		ϕ_1	$\{A, B\}$	
ϕ_2	$\{B, C\}$		ϕ_2	$\{A, C\}$	
ϕ_3	$\{C, D\}$		ϕ_3	$\{B, C\}$	
ϕ_4	$\{D, A\}$		ϕ_4	$\{C, D, E\}$	

But if we only know the graph $\mathcal{H}$, then the factorisation is **not unique**!

Can you tell why?
(Hint: by definition, the scope of factor is a complete subgraph)

But if we only know the graph $\mathcal{H}$, then the factorisation is **not unique**!

Can you tell why?
(Hint: by definition, the scope of factor is a complete subgraph)

Consider this graph

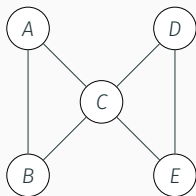One set of complete subgraphs for A, B, C, D:
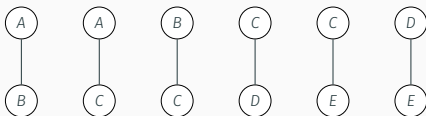
another set:

and combinations of these, and others...

There are many sets of complete subgraphs (cliques) that cover all nodes and edges in $\mathcal{H}$. Each set corresponds to a factorisation.

There are differently-factorised distributions that are consistent with $\mathcal{H}$.

This is different in BNs, whose DAG implies a unique factorisation.

Then the question to ask is: if $\mathcal{H}$ does not imply a factorisation, what then do all distributions that are 'compatible' with $\mathcal{H}$ have in common?

There are differently-factorised distributions that are consistent with $\mathcal{H}$.

This is different in BNs, whose DAG implies a unique factorisation.

Then the question to ask is: if $\mathcal{H}$ does not imply a factorisation, what then do all distributions that are 'compatible' with $\mathcal{H}$ have in common?

$\mathcal{H}$ implies a set of conditional independence statements and *that* is what all distributions that factorise over $\mathcal{H}$ have in common.

Summary

- An MN uses an undirected graph $\mathcal{H}$ to encode statements of (conditional) independence without ascribing directionality to interactions amongst rvs.

- To parameterise the MN, we specify a collection of factors (such as tabular factors) whose scopes are complete subgraphs.

- Unlike we did for BNs, we cannot read a unique factorisation readily from the MN structure.

- But, having a collection of factors, we can identify the graph. With the graph and its factors Φ, an MN is a complete representation of a joint distribution P_Φ.

LC3: MNs in code.

HC3ab: MNs – reasoning and influence.

WC3: exercises (semantics, reasoning and influence).

Reasoning

We begin by restating fundamental results of probability theory using general *factors* (as opposed to CPDs).

We then exploit those results and the MN representation to respond to queries about subsets of random variables and sets of observations.

We will start to see a connection between probabilistic inference and manipulation of the MN structure. This connection will be exploited more fully later in the course.

A Representation of our Uncertainty over A, B, C, D

With its undirected graph and factors, an MN identifies a *joint probability distribution* over the rvs of interest.
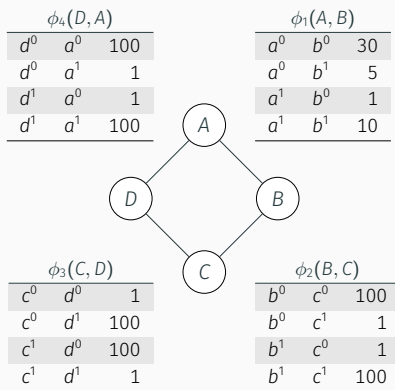
$\phi_4(D, A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_3(C, D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

The probability of any one of the possible outcomes of (A, B, C, D) in *Misconception* is proportional to $\phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A)$.

Probability Calculus

Marginalisation

$$P(A) = \sum_B P(A, B)$$

These results extend to the cases where *A* and/or *B* are sets of rvs.

Marginalisation

$$P(A) = \sum_B P(A, B) = \sum_B \frac{1}{Z}\tilde{P}(A, B)$$

These results extend to the cases where A and/or B are sets of rvs.

Marginalisation

$$P(A) = \sum_B P(A, B) = \sum_B \frac{1}{Z}\tilde{P}(A, B) = \frac{1}{Z}\underbrace{\sum_B \tilde{P}(A, B)}_{\tilde{P}(A)} \quad (2)$$

These results extend to the cases where *A* and/or *B* are sets of rvs.

Marginalisation

$$P(A) = \sum_B P(A, B) = \sum_B \frac{1}{Z}\tilde{P}(A, B) = \frac{1}{Z}\underbrace{\sum_B \tilde{P}(A, B)}_{\tilde{P}(A)} \qquad (2)$$

Conditioning

$$P(B|A) = \frac{P(A, B)}{P(A)}$$

These results extend to the cases where A and/or B are sets of rvs.

Marginalisation

$$P(A) = \sum_B P(A, B) = \sum_B \frac{1}{Z}\tilde{P}(A, B) = \frac{1}{Z}\underbrace{\sum_B \tilde{P}(A, B)}_{\tilde{P}(A)} \tag{2}$$

Conditioning

$$P(B|A) = \frac{P(A, B)}{P(A)} = \frac{\frac{1}{Z}\tilde{P}(A, B)}{\frac{1}{Z}\tilde{P}(A)}$$

These results extend to the cases where A and/or B are sets of rvs.

Marginalisation

$$P(A) = \sum_B P(A, B) = \sum_B \frac{1}{Z}\tilde{P}(A, B) = \frac{1}{Z}\underbrace{\sum_B \tilde{P}(A, B)}_{\tilde{P}(A)} \qquad (2)$$

Conditioning

$$P(B|A) = \frac{P(A, B)}{P(A)} = \frac{\frac{1}{Z}\tilde{P}(A, B)}{\frac{1}{Z}\tilde{P}(A)} = \frac{\tilde{P}(A, B)}{\tilde{P}(A)} \qquad (3)$$

These results extend to the cases where A and/or B are sets of rvs.

A	B	C	D	$\tilde{P}(A, B, C, D)$
a^0	b^0	c^0	d^0	300 000
a^0	b^0	c^0	d^1	300 000
a^0	b^0	c^1	d^0	300 000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5 000 000
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1 000 000
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100 000
a^1	b^1	c^1	d^0	100 000
a^1	b^1	c^1	d^1	100 000
Sum				7 201 840

1. $P(A = a^0) =$
2. $P(A = a^0, B = b^0) =$

Let's compute some marginals for the *Misconception* example

A	B	C	D	$\tilde{P}(A, B, C, D)$
a^0	b^0	c^0	d^0	300 000
a^0	b^0	c^0	d^1	300 000
a^0	b^0	c^1	d^0	300 000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5 000 000
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1 000 000
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100 000
a^1	b^1	c^1	d^0	100 000
a^1	b^1	c^1	d^1	100 000
Sum				7 201 840

1. $P(A = a^0) = \frac{\text{yellow}}{\text{blue}} = \frac{5901530}{7201840} \approx 0.8194$
2. $P(A = a^0, B = b^0) =$

Let's compute some marginals for the *Misconception* example

A	B	C	D	$\tilde{P}(A, B, C, D)$
a^0	b^0	c^0	d^0	300 000
a^0	b^0	c^0	d^1	300 000
a^0	b^0	c^1	d^0	300 000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5 000 000
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1 000 000
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100 000
a^1	b^1	c^1	d^0	100 000
a^1	b^1	c^1	d^1	100 000
	Sum			7 201 840

1. $P(A = a^0) = \frac{\text{yellow}}{\text{blue}} = \frac{5901530}{7201840} \approx 0.8194$
2. $P(A = a^0, B = b^0) =$

Let's compute some marginals for the *Misconception* example

A	B	C	D	$\tilde{P}(A, B, C, D)$
a^0	b^0	c^0	d^0	300 000
a^0	b^0	c^0	d^1	300 000
a^0	b^0	c^1	d^0	300 000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5 000 000
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1 000 000
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100 000
a^1	b^1	c^1	d^0	100 000
a^1	b^1	c^1	d^1	100 000
Sum				7 201 840

1. $P(A = a^0) = \frac{\text{yellow}}{\text{blue}} = \frac{5901530}{7201840} \approx 0.8194$
2. $P(A = a^0, B = b^0) = \frac{\text{dark yellow}}{\text{blue}} = \frac{900030}{7201840} \approx 0.1249$

A	B	C	D	$\tilde{P}(A, B, C, D)$
a^0	b^0	c^0	d^0	300 000
a^0	b^0	c^0	d^1	300 000
a^0	b^0	c^1	d^0	300 000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5 000 000
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1 000 000
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100 000
a^1	b^1	c^1	d^0	100 000
a^1	b^1	c^1	d^1	100 000
	Sum			7 201 840

1. $P(B = b^0 | A = a^0) =$
2. $P(B = b^1 | A = a^0) =$

Let's compute some conditionals for the *Misconception* example

A	B	C	D	$\tilde{P}(A, B, C, D)$
a^0	b^0	c^0	d^0	300 000
a^0	b^0	c^0	d^1	300 000
a^0	b^0	c^1	d^0	300 000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5 000 000
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1 000 000
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100 000
a^1	b^1	c^1	d^0	100 000
a^1	b^1	c^1	d^1	100 000
Sum				7 201 840

1. $P(B = b^0 | A = a^0) = \frac{\text{dark yellow}}{\text{light/dark yellow}} = \frac{900030}{5901530} \approx 0.1525$
2. $P(B = b^1 | A = a^0) = \frac{\text{light yellow}}{\text{light/dark yellow}} = \frac{900030}{5901530} \approx 0.8475$

Factor Operations

We can also reason using the MN representation, as opposed to the tabular view of the distribution.

For that, we need to learn more about how to manipulate factors:

- factor product
- factor normalisation
- factor reduction
- factor marginalisation   (Chapter 9)

Let $\phi_1(X, Y)$ and $\phi_2(Y, Z)$ be two factors.

The **factor product** $\phi_1(X, Y) \times \phi_2(Y, Z)$ is the factor ψ defines as:

$$\psi(X, Y, Z) = \phi_1(X, Y) \times \phi_2(Y, Z) \qquad (4)$$

The scope of the new factor is
$\text{Scope}[\psi] = \text{Scope}[\phi_1] \cup \text{Scope}[\phi_2] = \{X, Y, Z\}$.

This definition extends naturally to the case where X, Y, Z are
replaced by disjoint sets of rvs.

A	$\underline{B}$	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\times$

$\underline{B}$	C	ϕ_2
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

Dark green (b^1), light green (b^0).

A	$\underline{B}$	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\times$

$\underline{B}$	C	ϕ_2
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

=

A	B	C	$\psi_{A,B,C}$
a^0	b^0	c^0	3000
a^0	b^0	c^1	30
a^0	b^1	c^0	5
a^0	b^1	c^1	500
a^1	b^0	c^0	100
a^1	b^0	c^1	1
a^1	b^1	c^0	10
a^1	b^1	c^1	1000

Dark green (b^1), light green (b^0).

C	$\underline{D}$	ϕ_3
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\times$

$\underline{D}$	A	ϕ_4
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

Dark red (d^1), light red (d^0).

C	D	A	$\psi_{C,D,A}$
c^0	d^0	a^0	100
c^0	d^0	a^1	1
c^0	d^1	a^0	100
c^0	d^1	a^1	10000
c^1	d^0	a^0	10000
c^1	d^0	a^1	100
c^1	d^1	a^0	1
c^1	d^1	a^1	100

C	$\underline{D}$	ϕ_3
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\times$

$\underline{D}$	A	ϕ_4
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

=

Dark red (d^1), light red (d^0).

$\underline{A}$	B	$\underline{C}$	$\psi_{A,B,C}$
a^0	b^0	c^0	3000
a^0	b^0	c^1	30
a^0	b^1	c^0	5
a^0	b^1	c^1	500
a^1	b^0	c^0	100
a^1	b^0	c^1	1
a^1	b^1	c^0	10
a^1	b^1	c^1	1000

$\times$

$\underline{C}$	D	$\underline{A}$	$\psi_{C,D,A}$
c^0	d^0	a^0	100
c^0	d^0	a^1	1
c^0	d^1	a^0	100
c^0	d^1	a^1	10000
c^1	d^0	a^0	10000
c^1	d^0	a^1	100
c^1	d^1	a^0	1
c^1	d^1	a^1	100

Light yellow (a^0, c^0), dark yellow (a^0, c^1), light blue (a^1, c^0), dark blue (a^1, c^1).

$\underline{A}$	B	$\underline{C}$	$\psi_{A,B,C}$
a^0	b^0	c^0	3000
a^0	b^0	c^1	30
a^0	b^1	c^0	5
a^0	b^1	c^1	500
a^1	b^0	c^0	100
a^1	b^0	c^1	1
a^1	b^1	c^0	10
a^1	b^1	c^1	1000

$\times$

$\underline{C}$	D	$\underline{A}$	$\psi_{C,D,A}$
c^0	d^0	a^0	100
c^0	d^0	a^1	1
c^0	d^1	a^0	100
c^0	d^1	a^1	10000
c^1	d^0	a^0	10000
c^1	d^0	a^1	100
c^1	d^1	a^0	1
c^1	d^1	a^1	100

$=$

$\underline{A}$	B	$\underline{C}$	D	$\psi_{A,B,C,D}$
a^0	b^0	c^0	d^0	300 000
a^0	b^0	c^0	d^1	300 000
a^0	b^0	c^1	d^0	300 000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5×10^6
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1×10^6
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100 000
a^1	b^1	c^1	d^0	100 000
a^1	b^1	c^1	d^1	100 000

Light yellow (a^0, c^0), dark yellow (a^0, c^1), light blue (a^1, c^0), dark blue (a^1, c^1).

Let ϕ be a factor with scope X and $Z_\phi = \sum_{\mathsf{X}} \phi(\mathsf{X})$.

If $Z_\phi \in \mathbb{R}_{>0}$, we can obtain the normalised factor η:

$$\eta(\mathsf{X}) = \frac{1}{Z_\phi} \phi(\mathsf{X}) \tag{5}$$

A normalised factor may parameterise a joint probability distribution over X.

Factor Normalisation - *Misconception* Example

A	B	C	D	$\psi_{A,B,C,D} = \tilde{P}_\Phi$	$\eta_{A,B,C,D} = P_\Phi$
a^0	b^0	c^0	d^0	300 000	0.04
a^0	b^0	c^0	d^1	300 000	0.04
a^0	b^0	c^1	d^0	300 000	0.04
a^0	b^0	c^1	d^1	30	4.1×10^{-6}
a^0	b^1	c^0	d^0	500	6.9×10^{-5}
a^0	b^1	c^0	d^1	500	6.9×10^{-5}
a^0	b^1	c^1	d^0	5 000 000	0.69
a^0	b^1	c^1	d^1	500	6.9×10^{-5}
a^1	b^0	c^0	d^0	100	1.4×10^{-5}
a^1	b^0	c^0	d^1	1 000 000	0.14
a^1	b^0	c^1	d^0	100	1.4×10^{-5}
a^1	b^0	c^1	d^1	100	1.4×10^{-5}
a^1	b^1	c^0	d^0	10	1.4×10^{-6}
a^1	b^1	c^0	d^1	100 000	0.014
a^1	b^1	c^1	d^0	100 000	0.014
a^1	b^1	c^1	d^1	100 000	0.014
Sum				7 201 840	1.0

Factor Normalisation - Warning

While a normalised factor has the *form* of a joint distribution, not every normalised factor is coherent with a given Markov network!

For example, if we normalise $\psi_{A,B,C}$ the result 'looks' like a distribution over A, B, C:

A	B	C	$\psi_{A,B,C}$	$\eta_{A,B,C}$
a^0	b^0	c^0	3000	0.6458
a^0	b^0	c^1	30	0.0065
a^0	b^1	c^0	5	0.0011
a^0	b^1	c^1	500	0.1076
a^1	b^0	c^0	100	0.0215
a^1	b^0	c^1	1	0.0002
a^1	b^1	c^0	10	0.0021
a^1	b^1	c^1	1000	0.2152
Sum			4646	1.0

While $\eta_{A,B,C}$ *can* parameterise *some distribution* over A and C, **it is not the marginal distribution** $P_\Phi(A, C)$ of the Misconception example. Can you tell why?

While a normalised factor has the *form* of a joint distribution, not every normalised factor is coherent with a given Markov network!

For example, if we normalise $\psi_{A,B,C}$ the result 'looks' like a distribution over A, B, C:

A	B	C	$\psi_{A,B,C}$	$\eta_{A,B,C}$
a^0	b^0	c^0	3000	0.6458
a^0	b^0	c^1	30	0.0065
a^0	b^1	c^0	5	0.0011
a^0	b^1	c^1	500	0.1076
a^1	b^0	c^0	100	0.0215
a^1	b^0	c^1	1	0.0002
a^1	b^1	c^0	10	0.0021
a^1	b^1	c^1	1000	0.2152
	Sum		4646	1.0

While $\eta_{A,B,C}$ *can* parameterise *some distribution* over A and C, **it is not the marginal distribution** $P_\Phi(A, C)$ of the Misconception example. Can you tell why?

In the Misconception MN, A and C depend on D which is covered by factors ϕ_3 and ϕ_4 that have not been accounted for in $\psi_{A,B,C}$.

Let $\phi(X, Y)$ be a factor. We can marginalise a variable in the factor's scope. For example, we may create a new factor σ by summing over the values of Y as you would do for probability distributions:

$$\sigma(X) = \sum_{Y} \phi(X, Y) \tag{6}$$

The scope of the new factor is $\text{Scope}[\sigma] = \text{Scope}[\phi] \setminus \{Y\} = \{X\}$.

This definition extends naturally to the case where X, Y are replaced by disjoint sets of rvs.

Factor Marginalisation - *Misconception* $P_\Phi(A, B, C)$

A	B	C	$\underline{D}$	$\psi_{A,B,C,D}$
a^0	b^0	c^0	d^0	300000
a^0	b^0	c^0	d^1	300000
a^0	b^0	c^1	d^0	300000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5000000
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1000000
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100000
a^1	b^1	c^1	d^0	100000
a^1	b^1	c^1	d^1	100000

$\overset{\sum_D}{\Rightarrow}$

Colour: rows that differ only in their assignment of *D*.

A	B	C	$\underline{D}$	$\psi_{A,B,C,D}$
a^0	b^0	c^0	d^0	300000
a^0	b^0	c^0	d^1	300000
a^0	b^0	c^1	d^0	300000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5000000
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1000000
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100000
a^1	b^1	c^1	d^0	100000
a^1	b^1	c^1	d^1	100000

$$\overset{\sum_D}{\Longrightarrow}$$

A	B	C	$\sigma_{A,B,C}$
a^0	b^0	c^0	600000
a^0	b^0	c^1	300030
a^0	b^1	c^0	1000
a^0	b^1	c^1	5000500
a^1	b^0	c^0	1000100
a^1	b^0	c^1	200
a^1	b^1	c^0	100010
a^1	b^1	c^1	200000

$\sigma_{A,B,C}$ is $\tilde{P}(A, B, C)$

Colour: rows that differ only in their assignment of *D*.

Factor Marginalisation - *Misconception* $P_\Phi(A, B, C)$

A	B	C	$\underline{D}$	$\psi_{A,B,C,D}$
a^0	b^0	c^0	d^0	300000
a^0	b^0	c^0	d^1	300000
a^0	b^0	c^1	d^0	300000
a^0	b^0	c^1	d^1	30
a^0	b^1	c^0	d^0	500
a^0	b^1	c^0	d^1	500
a^0	b^1	c^1	d^0	5000000
a^0	b^1	c^1	d^1	500
a^1	b^0	c^0	d^0	100
a^1	b^0	c^0	d^1	1000000
a^1	b^0	c^1	d^0	100
a^1	b^0	c^1	d^1	100
a^1	b^1	c^0	d^0	10
a^1	b^1	c^0	d^1	100000
a^1	b^1	c^1	d^0	100000
a^1	b^1	c^1	d^1	100000

$\overset{\Sigma_D}{\Rightarrow}$

A	B	C	$\sigma_{A,B,C}$	$\eta_{A,B,C}$
a^0	b^0	c^0	600000	0.08331
a^0	b^0	c^1	300030	0.04166
a^0	b^1	c^0	1000	0.00014
a^0	b^1	c^1	5000500	0.69434
a^1	b^0	c^0	1000100	0.13887
a^1	b^0	c^1	200	0.00003
a^1	b^1	c^0	100010	0.01389
a^1	b^1	c^1	200000	0.02777

Normalising $\sigma_{A,B,C}$ gives us $P(A, B, C)$

Colour: rows that differ only in their assignment of *D*.

For a marginal factor to be coherent with a given MN, the factor's scope must account for all dependencies of the variables being marginalised.

A	$\underline{B}$	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\xrightarrow{\sum_B}$

Light brown (a^0), dark brown (a^1).

For a marginal factor to be coherent with a given MN, the factor's scope must account for all dependencies of the variables being marginalised.

A	$\underline{B}$	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\xrightarrow[B]{\sum}$

A	σ_A
a^0	35
a^1	11

This is not related to $\tilde{P}_\Phi(A)$ at all.

Light brown (a^0), dark brown (a^1).

For a marginal factor to be coherent with a given MN, the factor's scope must account for all dependencies of the variables being marginalised.

A	$\underline{B}$	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\overset{\sum_B}{\longrightarrow}$

A	σ_A
a^0	35
a^1	11

This is not related to $\tilde{P}_\Phi(A)$ at all.

How about $\sum_C \phi_1(A, B) \times \phi_2(B, C)$? Is that related to $\tilde{P}_\Phi(A, B)$?

Light brown (a^0), dark brown (a^1).

For a marginal factor to be coherent with a given MN, the factor's scope must account for all dependencies of the variables being marginalised.

A	B	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\overset{\sum_B}{\Longrightarrow}$

A	σ_A
a^0	35
a^1	11

This is not related to $\tilde{P}_\Phi(A)$ at all.

How about $\sum_C \phi_1(A, B) \times \phi_2(B, C)$? Is that related to $\tilde{P}_\Phi(A, B)$? No! Both *A* and *C* have dependencies not accounted for in the product.

Light brown (a^0), dark brown (a^1).

Let $\phi(X, Y)$ be a factor, and $Y = y$ an assignment for $Y \in \text{Scope}[\phi]$. We define **the reduction of the factor ϕ to the context $Y = y$**, denoted $\phi[Y = y]$ and abbreviated to $\phi[y]$ when unambiguous, to be the factor:

$$\phi[y](x) = \phi(x, y)$$

The scope of the new factor is $\text{Scope}[\phi] \setminus \{Y\} = \{X\}$.

This definition extends naturally to the case where X and or Y are replaced by disjoint sets of variables.

B	C	ϕ_2
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

$$\overset{C=c^1}{\Longrightarrow}$$

Dark pink (c^1), light pink (c^0).

B	C	ϕ_2
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

$$\underset{C=c^1}{\Rightarrow}$$

B	$\phi_2[c^1]$
b^0	1
b^1	100

Dark pink (c^1), light pink (c^0).

Reduced factors are factors, they can take part in factor products:

A	$\underline{B}$	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\times$

$\underline{B}$	$\phi_2[c^1]$
b^0	1
b^1	100

Dark green (b^1), light green (b^0).

Reduced factors are factors, they can take part in factor products:

A	$\underline{B}$	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\times$

$\underline{B}$	$\phi_2[c^1]$
b^0	1
b^1	100

$=$

A	B	$\psi_{A,B,C}[c^1]$
a^0	b^0	30
a^0	b^1	500
a^1	b^0	1
a^1	b^1	1000

Dark green (b^1), light green (b^0).

Reduced factors are factors, they can take part in factor products:

A	B	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\times$

B	$\phi_2[c^1]$
b^0	1
b^1	100

$=$

A	B	$\psi_{A,B,C}[c^1]$
a^0	b^0	30
a^0	b^1	500
a^1	b^0	1
a^1	b^1	1000

Do you think this is true?

$$(\phi_1(A, B) \times \phi_2(B, C))[c^1] = \phi_1(A, B) \times \phi_2[c^1](B)$$

Dark green (b^1), light green (b^0).

Reduced factors are factors, they can take part in factor products:

A	$\underline{B}$	ϕ_1
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\times$

$\underline{B}$	$\phi_2[c^1]$
b^0	1
b^1	100

$=$

A	B	$\psi_{A,B,C}[c^1]$
a^0	b^0	30
a^0	b^1	500
a^1	b^0	1
a^1	b^1	1000

Do you think this is true?

$$(\phi_1(A, B) \times \phi_2(B, C))[c^1] = \phi_1(A, B) \times \phi_2[c^1](B)$$

Yes! And, if we are working with tabular factors, the second option is more efficient since we do not have to create the large table $\psi_{A,B,C}$.

Dark green (b^1), light green (b^0).

Let's clear out a potential confusion about factors and their scopes.

The scope of a general factor is any collection of rvs.

But to *factorise* $P_\Phi(X_1, \ldots, X_n)$ according to the MN structure $\mathcal{H}$, the collection of factors Φ must have cliques as their scopes. This constraint is specific to the factors in the collection Φ.

Once we start manipulating factors to characterise other views of $\tilde{P}_\Phi$ and P_Φ, we will encounter factors whose scopes are not cliques.

Example:
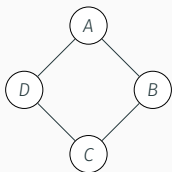
Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

For $\phi_i \in \Phi$, $\mathbf{D}_i = \text{Scope}[\phi_i]$ is a clique of $\mathcal{H}$. Other factors are not constrained to cliques:

- $\tilde{P}(A, B, C, D) = \prod_{\phi_i \in \Phi} \phi_i(\mathbf{D}_i)$

- $\tilde{P}(A, C) = \sum_{B,D} \prod_{\phi_i \in \Phi} \phi_i(\mathbf{D}_i)$

- $\psi(A, B, C) = \phi_1(A, B)\phi_2(B, C)$, etc.

If we find that Debbie and Bob have the misconception, that is, $(D = d^1, B = b^1)$, then we can obtain a factor that is proportional to $P_\Phi(A, C | B = b^1, D = d^1)$ via:

$$(\phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A))[b^1, d^1]$$

which is $\tilde{P}_\Phi(A, B = b^1, C, D = d^1)$.

And, of course, $P_\Phi(A, C | B = b^1, D = d^1) \propto \tilde{P}_\Phi(A, B = b^1, C, D = d^1)$

If we find that Debbie and Bob have the misconception, that is, $(D = d^1, B = b^1)$, then we can obtain a factor that is proportional to $P_\Phi(A, C | B = b^1, D = d^1)$ via:

$$(\phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A))[b^1, d^1]$$

which is $\tilde{P}_\Phi(A, B = b^1, C, D = d^1)$.

And, of course, $P_\Phi(A, C | B = b^1, D = d^1) \propto \tilde{P}_\Phi(A, B = b^1, C, D = d^1)$

Alternatively, we can reduce the factors first:

$$\psi_{b1d1}(A, C) = \phi_1[b^1](A) \times \phi_2[b^1](C) \times \phi_3[d^1](C) \times \phi_4[d^1](A)$$

Then, to obtain the distribution, we simply normalise ψ_{b1d1}.

Exercise

Express $P_\Phi(A, C | B = b^1, D = d^1)$ for the Misconception MN.

_____

Shaded columns are shown as a reminder of what the original factor was before we reduced it.

43

Exercise

Express $P_\Phi(A, C | B = b^1, D = d^1)$ for the Misconception MN.

A	B	$\phi_1[b_1]$
a^0	b^1	5
a^1	b^1	10

$\times$

B	C	$\phi_2[b^1]$
b^1	c^0	1
b^1	c^1	100

$\times$

C	D	$\phi_3[d^1]$
c^0	d^1	100
c^1	d^1	1

$\times$

D	A	$\phi_4[d^1]$
d^1	a^0	1
d^1	a^1	100

=

A	C	$\tilde{P}_\Phi[b^1, d^1]$	$P_\Phi(A, C | B = b^1, D = d^1)$
a^0	c^0		
a^0	c^1		
a^1	c^0		
a^1	c^1		
Sum			

Shaded columns are shown as a reminder of what the original factor was before we reduced it.

Exercise

Express $P_\Phi(A, C | B = b^1, D = d^1)$ for the Misconception MN.

A	B	$\phi_1[b_1]$
a^0	b^1	5
a^1	b^1	10

$\times$

B	C	$\phi_2[b^1]$
b^1	c^0	1
b^1	c^1	100

$\times$

C	D	$\phi_3[d^1]$
c^0	d^1	100
c^1	d^1	1

$\times$

D	A	$\phi_4[d^1]$
d^1	a^0	1
d^1	a^1	100

=

A	C	$\tilde{P}_\Phi[b^1, d^1]$	$P_\Phi(A, C | B = b^1, D = d^1)$
a^0	c^0	500	0.0025
a^0	c^1	500	0.0025
a^1	c^0	100000	0.4975
a^1	c^1	100000	0.4975
Sum		201000	1.0

Shaded columns are shown as a reminder of what the original factor was before we reduced it.

Exercise

Express $P_\Phi(A, C|B = b^1, D = d^1)$ for the Misconception MN.

A	B	$\phi_1[b_1]$		B	C	$\phi_2[b^1]$		C	D	$\phi_3[d^1]$		D	A	$\phi_4[d^1]$
a^0	b^1	5	×	b^1	c^0	1	×	c^0	d^1	100	×	d^1	a^0	1
a^1	b^1	10		b^1	c^1	100		c^1	d^1	1		d^1	a^1	100

	A	C	$\tilde{P}_\Phi[b^1, d^1]$	$P_\Phi(A, C|B = b^1, D = d^1)$
	a^0	c^0	500	0.0025
=	a^0	c^1	500	0.0025
	a^1	c^0	100000	0.4975
	a^1	c^1	100000	0.4975
	Sum		201000	1.0

Can you now assess the marginal probability $P(B = b^1, D = d^1)$?

Shaded columns are shown as a reminder of what the original factor was before we reduced it.

Exercise

Express $P_\Phi(A, C | B = b^1, D = d^1)$ for the Misconception MN.

A	B	$\phi_1[b_1]$
a^0	b^1	5
a^1	b^1	10

$\times$

B	C	$\phi_2[b^1]$
b^1	c^0	1
b^1	c^1	100

$\times$

C	D	$\phi_3[d^1]$
c^0	d^1	100
c^1	d^1	1

$\times$

D	A	$\phi_4[d^1]$
d^1	a^0	1
d^1	a^1	100

=

A	C	$\tilde{P}_\Phi[b^1, d^1]$	$P_\Phi(A, C | B = b^1, D = d^1)$
a^0	c^0	500	0.0025
a^0	c^1	500	0.0025
a^1	c^0	100000	0.4975
a^1	c^1	100000	0.4975
Sum		201000	1.0

Can you now assess the marginal probability $P(B = b^1, D = d^1)$? It is $\frac{201000}{Z}$, we still need the normaliser Z of the the complete $\tilde{P}_\Phi$.

Shaded columns are shown as a reminder of what the original factor was before we reduced it.

Summary

An MN identifies a joint distribution, hence it is possible to reason marginally or conditionally.

Reasoning can be formulated in terms of operations on factors:

- Product: produces larger factors coherent with the MN
  use: build the joint measure
- Reduction: produces smaller factors coherent with the MN
  use: assign rvs and/or condition on evidence
- Normalisation (without missing dependencies): produces a normalised measure
  use: turn a view of $\tilde{P}_\Phi$ into P_Φ
- Marginalisation (without missing dependencies): produces smaller factors coherent with marginals of the MN
  use: marginalise rvs in a view of $\tilde{P}_\Phi$

We will turn to influence and independence.

Influence

We revisit independence.

Then trace influence in MNs to ascertain graphical separation.

We conclude with an overview of the complete set of conditional independences implied by the MN structure.

Independence (revisited)

A joint distribution *P* **satisfies independence between** the random variables *X* and *Y*, which we denote by $P \models A \perp B$, if:

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

A joint distribution *P* **satisfies independence** of the rvs *X* and *Y* *given Z*, which we denote by $P \models (X \perp Y \mid Z)$, if:

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

Independence (revisited)

A joint distribution P **satisfies independence between** the random variables X and Y, which we denote by $P \models A \perp B$, if:

- $P(X, Y) = P(X)P(Y)$
- $P(X|Y) = P(X)$
- $P(Y|X) = P(Y)$
- $P(X, Y) \propto \phi_1(X)\phi_2(Y)$

A joint distribution P **satisfies independence** of the rvs X and Y *given* Z, which we denote by $P \models (X \perp Y \mid Z)$, if:

- $P(X, Y|Z) = P(X|Z)P(Y|Z)$
- $P(X|Y, Z) = P(X|Z)$
- $P(Y|X, Z) = P(Y|Z)$
- $P(X, Y, Z) \propto \phi_1(X, Z)\phi_2(Y, Z)$

where ϕ_1 and ϕ_2 are so-called **factors**.

Independence

Does $A \perp C | B, D$ hold?

B	D	A	C	$\tilde{P}(A, B, C, D)$
b^0	d^0	a^0	c^0	300 000
b^0	d^0	a^0	c^1	300 000
b^0	d^0	a^1	c^0	100
b^0	d^0	a^1	c^1	100
b^0	d^1	a^0	c^0	300 000
b^0	d^1	a^0	c^1	30
b^0	d^1	a^1	c^0	1×10^6
b^0	d^1	a^1	c^1	100
b^1	d^0	a^0	c^0	500
b^1	d^0	a^0	c^1	5×10^6
b^1	d^0	a^1	c^0	10
b^1	d^0	a^1	c^1	100 000
b^1	d^1	a^0	c^0	500
b^1	d^1	a^0	c^1	500
b^1	d^1	a^1	c^0	100 000
b^1	d^1	a^1	c^1	100 000

Colour: (B, D); dark: a^0; light: a^1.

Does $B \perp C | A, D$ hold?

A	D	B	C	$\tilde{P}(A, B, C, D)$
a^0	d^0	b^0	c^0	300 000
a^0	d^0	b^0	c^1	300 000
a^0	d^0	b^1	c^0	500
a^0	d^0	b^1	c^1	5×10^6
a^0	d^1	b^0	c^0	300 000
a^0	d^1	b^0	c^1	30
a^0	d^1	b^1	c^0	500
a^0	d^1	b^1	c^1	500
a^1	d^0	b^0	c^0	100
a^1	d^0	b^0	c^1	100
a^1	d^0	b^1	c^0	10
a^1	d^0	b^1	c^1	100 000
a^1	d^1	b^0	c^0	1×10^6
a^1	d^1	b^0	c^1	100
a^1	d^1	b^1	c^0	100 000
a^1	d^1	b^1	c^1	100 000

Colour: (A, D); dark: b^0; light: b^1.

Independence

Does $A \perp C | B, D$ hold? ✓

B	D	A	C	$\tilde{P}(A,B,C,D)$
b^0	d^0	a^0	c^0	300 000
b^0	d^0	a^0	c^1	300 000
b^0	d^0	a^1	c^0	100
b^0	d^0	a^1	c^1	100
b^0	d^1	a^0	c^0	300 000
b^0	d^1	a^0	c^1	30
b^0	d^1	a^1	c^0	1×10^6
b^0	d^1	a^1	c^1	100
b^1	d^0	a^0	c^0	500
b^1	d^0	a^0	c^1	5×10^6
b^1	d^0	a^1	c^0	10
b^1	d^0	a^1	c^1	100 000
b^1	d^1	a^0	c^0	500
b^1	d^1	a^0	c^1	500
b^1	d^1	a^1	c^0	100 000
b^1	d^1	a^1	c^1	100 000

Colour: (B, D); dark: a^0; light: a^1.

Does $B \perp C | A, D$ hold? ✗

A	D	B	C	$\tilde{P}(A,B,C,D)$
a^0	d^0	b^0	c^0	300 000
a^0	d^0	b^0	c^1	300 000
a^0	d^0	b^1	c^0	500
a^0	d^0	b^1	c^1	5×10^6
a^0	d^1	b^0	c^0	300 000
a^0	d^1	b^0	c^1	30
a^0	d^1	b^1	c^0	500
a^0	d^1	b^1	c^1	500
a^1	d^0	b^0	c^0	100
a^1	d^0	b^0	c^1	100
a^1	d^0	b^1	c^0	10
a^1	d^0	b^1	c^1	100 000
a^1	d^1	b^0	c^0	1×10^6
a^1	d^1	b^0	c^1	100
a^1	d^1	b^1	c^0	100 000
a^1	d^1	b^1	c^1	100 000

Colour: (A, D); dark: b^0; light: b^1.

Direct Influence

If there's an edge connecting *X* and *Y* in $\mathcal{H}$, then these variables will influence one another in any *P* factorising over $\mathcal{H}$.

Rationale:

- If there is an edge, for any P_Φ factorising over $\mathcal{H}$, there must be at least one factor ϕ covering this edge. The value of this factor will clearly depend on both *X* and *Y*.
  If it didn't, the factor would not 'cover' the edge.

This is similar to how it worked in BNs.

As edges are undirected, a node *X* can influence a node *Y* so long as there is a path connecting them.

Let's see this by characterising $P_\Phi(A, C)$ in the Misconception example:

$$P_\Phi(A, C) \propto \sum_B \sum_D \phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A)$$

Reminder: $\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A)$

As edges are undirected, a node *X* can influence a node *Y* so long as there is a path connecting them.

Let's see this by characterising $P_\Phi(A, C)$ in the Misconception example:

$$P_\Phi(A, C) \propto \sum_B \sum_D \phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A)$$

$$\propto \sum_B \phi_1(A, B) \times \phi_2(B, C) \times \sum_D \phi_3(C, D) \times \phi_4(D, A)$$

See that because of the sums, we cannot re-express this as two factors $\psi_1(A) \times \psi_2(C)$, hence $P_\Phi \not\models A \perp C$.

Reminder: $\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A)$

If the paths connecting *X* and *Y* in $\mathcal{H}$ are 'blocked' by observations, then *X* and *Y* cannot influence one another.

To see this, let's use an example: can *A* influence *D* in the Misconception example when we observe $B = b$ and $D = d$?

$$P_\Phi(A, B = b, C, D = d) \propto \phi_1[b](A) \times \phi_2[b](C) \times \phi_3[d](C) \times \phi_4[d](A)$$

_____

Reminder: $\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A)$

If the paths connecting *X* and *Y* in $\mathcal{H}$ are 'blocked' by observations, then *X* and *Y* cannot influence one another.

To see this, let's use an example: can *A* influence *D* in the Misconception example when we observe $B = b$ and $D = d$?

$$P_\Phi(A, B = b, C, D = d) \propto \phi_1[b](A) \times \phi_2[b](C) \times \phi_3[d](C) \times \phi_4[d](A)$$

$$\propto \underbrace{\phi_1[b](A) \times \phi_4[d](A)}_{\psi_1(A)} \times \underbrace{\phi_2[b](C) \times \phi_3[d](C)}_{\psi_2(C)}$$

Reminder: $\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A)$

If the paths connecting *X* and *Y* in $\mathcal{H}$ are 'blocked' by observations, then *X* and *Y* cannot influence one another.

To see this, let's use an example: can *A* influence *D* in the Misconception example when we observe $B = b$ and $D = d$?

$$P_\Phi(A, B = b, C, D = d) \propto \phi_1[b](A) \times \phi_2[b](C) \times \phi_3[d](C) \times \phi_4[d](A)$$
$$\propto \underbrace{\phi_1[b](A) \times \phi_4[d](A)}_{\psi_1(A)} \times \underbrace{\phi_2[b](C) \times \phi_3[d](C)}_{\psi_2(C)}$$

See that with factors reduced by observations, we can rearrange the terms as two independent factors $\psi_1(A) \times \psi_2(C)$, hence $P_\Phi \models A \perp C | B, D$.

Reminder: $\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B) \times \phi_2(B, C) \times \phi_3(C, D) \times \phi_4(D, A)$

A path $X_1, \ldots, X_k$ in $\mathcal{H}$ is **active** *given* Z if no X_i in the path is in Z.

A set of nodes Z separates X and Y in $\mathcal{H}$, denoted sep(X; Y|Z), if there is no active path between any node $X \in$ X and $Y \in$ Y given Z.

As in the case of BNs, graphical separation implies conditional independence and vice-versa. But now we have *separation* rather than *directed-separation*.
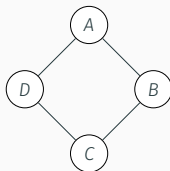
Any distribution P that factorises over the MN structure $\mathcal{H}$ satisfies the **global independencies**:

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

A node is independent of all other nodes given its immediate neighbours. For example, $A \perp C \mid B, D$

The set of immediate neighbours of a node is called its *Markov Blanket*, denoted $\text{MB}_\mathcal{H}(X_i)$.

$$\mathcal{I}_l(\mathcal{H}) = \{X_i \perp \text{Rest}_\mathcal{H}(X_i) \mid \text{MB}_\mathcal{H}(X_i) : \text{for all } X_i\}$$

When a distribution P that factorises over $\mathcal{H}$ is 'positive' (no zero probabilities allowed), the local independencies $\mathcal{I}_l(\mathcal{H})$ hold and imply the global ones.

The 'rest' of the nodes in $\mathcal{H}$ relative to X_i is $\mathcal{X} \setminus \{X_i\}$, we denote this by $\text{Rest}_\mathcal{H}(X_i)$. The textbook does not use this (much cleaner) notation.
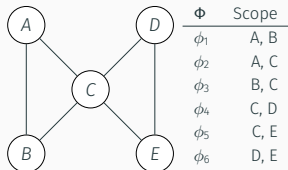
We warned you against marginalisating and/or normalising a factor whose rvs in scope have dependencies that are 'unnacounted for'.

Now we can make that even more concrete.

It is safe to use these operations on rvs that separated from the rest.

Consider this MN

Φ	Scope
ϕ_1	A, B
ϕ_2	A, C
ϕ_3	B, C
ϕ_4	C, D
ϕ_5	C, E
ϕ_6	D, E

We warned you **against marginalisating and/or normalising a factor whose rvs in scope have dependencies that are 'unnacounted for'**.

Now we can make that even more concrete.

It is safe to use these operations on rvs that separated from the rest.

Consider this MN

Φ	Scope
ϕ_1	A, B
ϕ_2	A, C
ϕ_3	B, C
ϕ_4	C, D
ϕ_5	C, E
ϕ_6	D, E
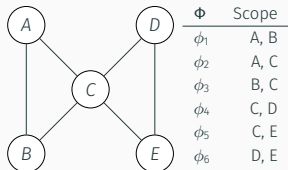
Observing $C = c$, separates (A, B) and (D, E). So, we work with $(A, B, C = c)$ ignoring (D, E), or vice-versa.

Example:
$$\tilde{P}_\Phi(A, B | C = c) = \phi_1(A, B) \times \phi_2[c](A) \times \phi_3[c](B)$$
We may marginalise/normalise!

Summary

- Independence can be tested using factors.
- The MN structure (no need to know the factors) is sufficient to ascertain conditional independences, via graphical separation.
- While the MN structure does not identify a unique factorisation, it identifies a unique set of global independencies.
- Separation allows us to characterise this set.
- The MN counterpart to local independencies uses the notion of Markov blanket (immediate neighbours).[20]

---

[20]There are some formal analysis tools that break in the special case of distributions that allow for 0 probabilities, but we will avoid such distributions in this course.

WC3: exercises (semantics, reasoning and influence).

Friday: P3 deadline.

Midterm: BNs and MNs.

After midterm: Inference.

References

[1] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.