Markov Networks

Or undirected graphical models

Wilker Aziz

w.aziz@uva.nl

Fall 2025 (v1)

https://probabll.github.io

Outline and goals

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].

Overview of Module 2

HC2b: MNs – semantics.

HC3a: MNs – reasoning.

LC3: MNs in code.

HC3b: MNs – influence.

WC3: exercises (semantics, reasoning and influence).

Table of contents

Semantics

Outline for this section

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.

The *Misconception* example (Example 3.8)

Four students (**A**lice, **B**ob, **C**harles and **D**ebbie) 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.

Conditional Independences

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

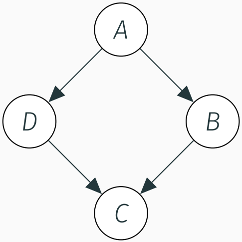
Can we represent these two statements with a BN?

Conditional Independences

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:

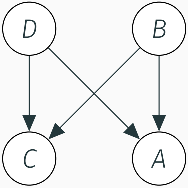
- $A \perp C \mid B, D$ ✓
- but D, B are independent given A alone ✗
- and if we are given A, C then D, B become dependent ✗

Conditional Independences

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

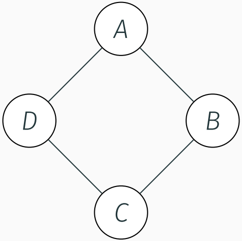


Figure 1: The MN graph for the *Misconception* example

We need an **undirected** representation of the relevant interactions!

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

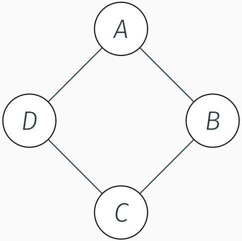
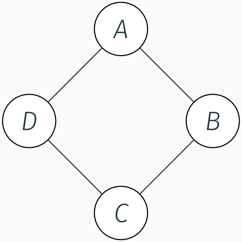


Figure 1: The MN graph for the *Misconception* example

We need an **undirected** representation of the relevant interactions! But how can we parameterise a joint distribution coherent with this view?

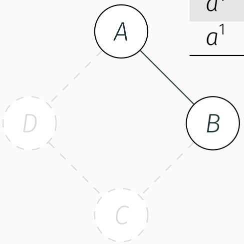
Factors for the *Misconception* example (Fig 4.1 of textbook)

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



Factors for the *Misconception* example (Fig 4.1 of textbook)

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

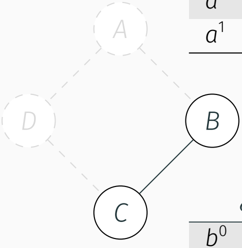


$\phi_1(A, B)$

a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

Factors for the *Misconception* example (Fig 4.1 of textbook)

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

$\phi_2(B, C)$

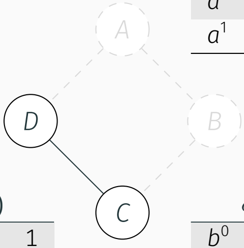
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

Factors for the *Misconception* example (Fig 4.1 of textbook)

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

Factors for the *Misconception* example (Fig 4.1 of textbook)

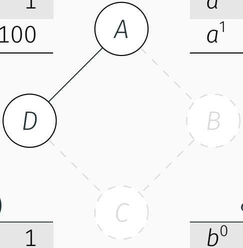
A factor ϕ with scope X, Y represents the undirected *affinity* between possible outcomes of X and Y.

$\phi_4(D, A)$

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

Factors for the *Misconception* example (Fig 4.1 of textbook)

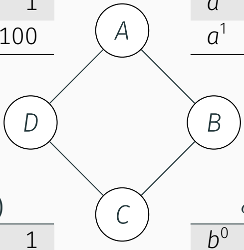
A factor ϕ with scope X, Y represents the undirected *affinity* between possible outcomes of X and Y.

$\phi_4(D, A)$

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

Factor

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$[\phi]$.

Factor

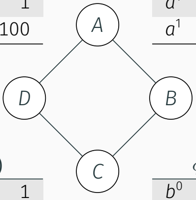
A factor ϕ is a function from the joint outcome space of a set D of rvs, which we call the factor's *scope*, to $\mathbb{R}_{\geq 0}$. Factors represent affinity amongst interacting variables.

$\phi_4(D, A)$

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$[\phi]$.

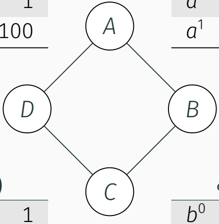
How can we define these joint probabilities?

$\phi_4(D, A)$

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

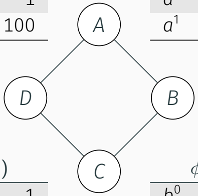
A *joint probability distribution* is a means to represent our uncertainty over the *possible* assignments of the rvs of interest.

$\phi_4(D, A)$

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



$\phi_3(C, D)$

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

With its undirected graph and factors, an MN identifies one such distribution.

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

Joint Distribution for the Misconception Example

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 $\mathrm{Val}(A) \times \mathrm{Val}(B) \times \mathrm{Val}(C) \times \mathrm{Val}(D)$

Compactness

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

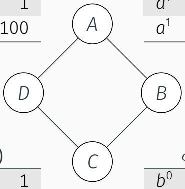
The probability value of any one of the joint outcomes is expressed via a product of local factors for directly interacting variables.

$\phi_4(D, A)$

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^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

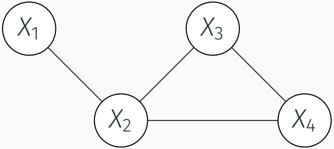
$\phi_2(B, C)$

b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

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.

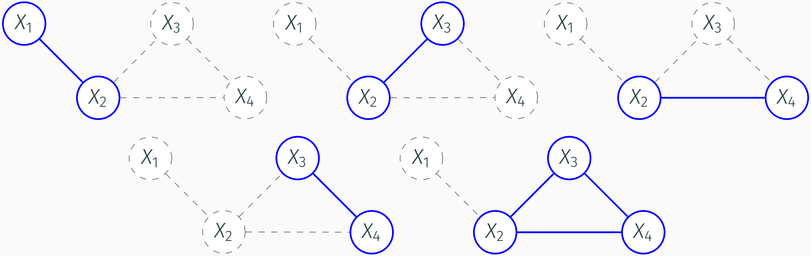
While in this example (K=2, N=4), K^N happens to be equal to NK^2, for larger K and N the MN is very compact.

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.



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

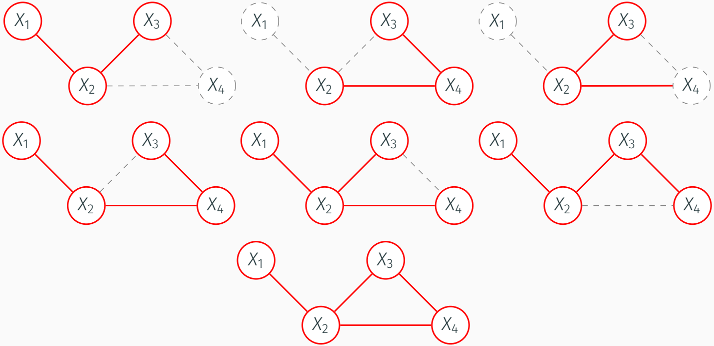
The notion of *directly interacting variables* corresponds to what is known as **complete subgraphs**. The factors in the MN, by definition, have complete subgraphs as their scope.



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

Complete Subgraphs (definition 2.13)

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.



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

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 rvs in the graph (i.e., $\cup_i \mathbf{D}_i$ is $\{X_1, \ldots, X_n\}$).

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 D_i of a factor $\phi_i \in \Phi$ is a complete subgraph of $\mathcal{H}$; Φ covers all rvs in the graph (i.e., $\cup_i \mathsf{D}_i$ is $\{X_1, \ldots, X_n\}$).

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

$$P(X_1, \ldots, X_n) \stackrel{\mathcal{H},\Phi}{=} \frac{1}{Z} \prod_{\phi_i \in \Phi} \phi_i(\mathsf{D}_i) \tag{1}$$

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

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

Graphs and Distributions (1/3)

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

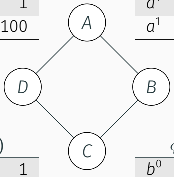
In the *Misconception* example:

$\phi_4(D, A)$

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.

Graphs and Distributions (2/3)

If we only have the graph $\mathcal{H}$, then the factorisation is **not unique**!

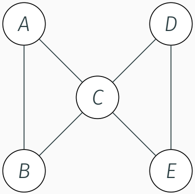
Can you tell why?
(Hint: by definition, the scope of factor is a complete subgraph)

Graphs and Distributions (2/3)

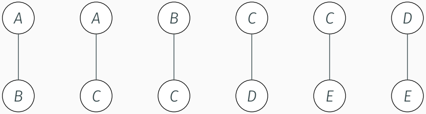
If we only have the graph $\mathcal{H}$, then the factorisation is **not unique**!

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

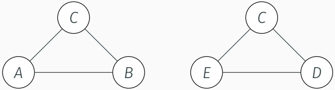
Consider this graph



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 the nodes A, B, C, D, E. Each set corresponds to a factorisation.

Graphs and Distributions (3/3)

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.

Global Independencies $\mathcal{I}(\mathcal{H})$

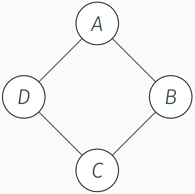
A set of nodes $\mathbf{Z}$ separates $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{H}$, denoted $\mathrm{sep}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$, if there is no active path between any node $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ given $\mathbf{Z}$.

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

If a distribution P factorises over the MN structure $\mathcal{H}$, then the global independencies $\mathcal{I}(\mathcal{H})$ must hold.

Local Independencies $\mathcal{I}_l(\mathcal{H})$

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 $\mathrm{MB}_{\mathcal{H}}(X_i)$.

$$\mathcal{I}_l(\mathcal{H}) = \{X_i \perp \mathcal{X} - \{X_i\} - \mathrm{MB}_{\mathcal{H}}(X_i) | \mathrm{MB}_{\mathcal{H}}(X_i) : \text{for all } X_i\}$$

If a distribution P factorises over the MN structure $\mathcal{H}$, then the global independencies $\mathcal{I}(\mathcal{H})$ must hold. If P is positive (no zero probabilities allowed), the local independencies $\mathcal{I}_l(\mathcal{H})$ hold and imply the global ones.

Summary

- An MN uses an undirected graph $\mathcal{H}$ to encode statements of (conditional) independence, and these must hold in any distribution P that factorises over $\mathcal{H}$.
- Unlike we did for BNs, we cannot read a unique factorisation readily from the MN structure: because there are numerous different sets of complete subgraphs that cover all of the rvs.
- To each complete subgraph, an MN may associate a factor (such as a tabular factor). Not all complete subgraphs need factors, so long as all rvs are 'covered'.
- With a collection of factors Φ that covers all rvs, an MN is a complete representation of a joint distribution P.

What’s Next?

LC3: MNs in code.

HC3ab: MNs – reasoning and influence.

WC3: exercises (semantics, reasoning and influence).

Reasoning

Outline for this section

We begin by restating three of the most fundamental results of probability theory.

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

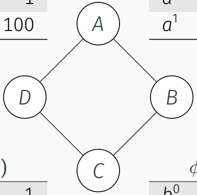
With its undirected graph and factors, an MN identifies a *joint probability distribution* over the rvs of interest.

$\phi_4(D, A)$

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.

Probability Calculus

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.

Probability Calculus

Marginalisation

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

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

Probability Calculus

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

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

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

Probability Calculus

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.

Probability Calculus

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

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

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

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

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

Summary

- An MN identifies a joint distribution, hence it is possible to reason marginally or conditionally.
- We can perform marginalisation and conditioning using the unnormalised measure
    - but normalising the marginals still requires Z,
    - and for conditionals Z cancels out.
- Later in the course we will develop a general graph-based algorithm so we can perform these operations without necessarily constructing the tabular view of the distribution.

What Next?

We will turn to influence and independence.

References

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