Chapter 2:

Preliminaries

Random variables

Let $\boldsymbol{V} = \{V_1, \ldots, V_n\}$, $n \geq 1$, be a set of random variables. Each variable $V_i \in \boldsymbol{V}$ can take on one of $m \geq 2$ values; for now we consider 2-valued variables:

- $V_i = true$, denoted by v_i;
- $V_i = false$, denoted by $\neg v_i$ (or by $\overline{v_i}$).

The set $\boldsymbol{V}$ spans a Boolean Algebra of logical propositions $\mathcal{V}$:

- T(rue), F(alse) $\in \mathcal{V}$;
- for all variables $V_i \in \boldsymbol{V}$ we have that $v_i \in \mathcal{V}$;
- for all $x \in \mathcal{V}$ we have that $\neg x \in \mathcal{V}$;
- for all $x, y \in \mathcal{V}$ we have that $x \wedge y \in \mathcal{V}$ and $x \vee y \in \mathcal{V}$.

The elements of $\mathcal{V}$ obey the usual rules of propositional logic.

The joint probability distribution

Definition:

Let $\mathcal{V}$ be the Boolean Algebra of propositions spanned by a set of random variables $\boldsymbol{V}$. Let $\Pr : \mathcal{V} \to [0, 1]$ be a function such that

- $\Pr$ is positive: for each $x \in \mathcal{V}$ we have that $\Pr(x) \geq 0$ and, more specifically, $\Pr(\mathsf{F}) = 0$;
- $\Pr$ is normed: $\Pr(\mathsf{T}) = 1$;
- $\Pr$ is additive: we have, for each $x, y \in \mathcal{V}$ with $x \wedge y \equiv \mathsf{F}$, that $\Pr(x \vee y) = \Pr(x) + \Pr(y)$.

The function $\Pr$ is a joint probability distribution on $\boldsymbol{V}$; the function value $\Pr(x)$ is the probability of x.

Independence of propositions

Definition: Let $\mathcal{V}$ be the Boolean Algebra of propositions spanned by a set of random variables $\boldsymbol{V}$. Let $\Pr$ be a joint probability distribution on V.

Two propositions $x, y \in \mathcal{V}$ are called independent in $\Pr$ if

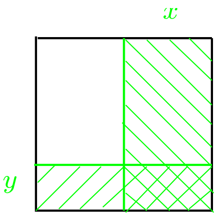
$$\Pr(x \wedge y) = \Pr(x) \cdot \Pr(y)$$

The propositions $x, y \in \mathcal{V}$ are called conditionally independent given the proposition $z \in \mathcal{V}$ if we have that

$$\Pr(x \wedge y \mid z) = \Pr(x \mid z) \cdot \Pr(y \mid z)$$

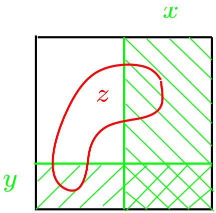
The two notions of independence (1)

- Consider two propositions $x, y \in \mathcal{V}$ such that x and y are independent [1]:



Can $z \in \mathcal{V}$ exist such that x and y are dependent given z?

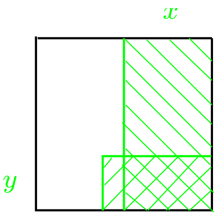
- Yes:



[1] The square has area 1, representing the total probability mass.

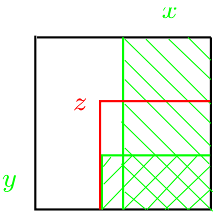
The two notions of independence (2)

- Consider two propositions $x, y \in \mathcal{V}$ such that x and y are dependent:



  Can $z \in \mathcal{V}$ exist such that x and y are conditionally independent given z?

- Yes:



Configurations

Let $\boldsymbol{V}$ be a set of random variables and let $\boldsymbol{W} \subseteq \boldsymbol{V}$.

- a configuration $c_{\boldsymbol{W}}$ of $\boldsymbol{W}$ is a conjunction of value assignments to the variables from $\boldsymbol{W}$;
- convention: $c_{\emptyset} =$ T;
- $\boldsymbol{w}$ is used to denote a specific configuration of $\boldsymbol{W}$.
- $\boldsymbol{W}$ also indicates all possible configurations to the set $\boldsymbol{W}$ (notation abuse!): *$\boldsymbol{W}$ is then considered to be a template that can be filled in with any configuration $c_{\boldsymbol{W}}$.*

Example: Let $\boldsymbol{W} = \{V_1, V_3, V_7\}$. $\boldsymbol{W} = V_1 \wedge V_3 \wedge V_7$ denotes a configuration template: filling in values for V_i results in proper propositions/configurations. Some configurations $c_{\boldsymbol{W}}$ of $\boldsymbol{W}$ are:

$$\begin{array}{rcrcr} V_1 = \mathit{true} & \wedge & V_3 = \mathit{true} & \wedge & V_7 = \mathit{false} \\ v_1 & \wedge & \neg v_3 & \wedge & v_7 \\ \neg v_1 & \wedge & v_3 & \wedge & \neg v_7 \end{array}$$

■

Conventions and notation

In the remainder of this course, for distributions on $\boldsymbol{V}$:

- rather than talking about propositions $x \in \mathcal{V}$ spanned by $\boldsymbol{V}$ we refer to configurations $c_{\boldsymbol{V}}$ of $\boldsymbol{V}$

	Set (bold faced)	Singleton
Variables/templates (capital)	$\boldsymbol{V}$	V
Values/configurations	$c_{\boldsymbol{V}}$, $\boldsymbol{v}$	c_V, v

- conjunctions are often left implicit: e.g. $v_1\, v_2$ denotes $v_1 \wedge v_2$;
- note the following differences (!)
  - probabilities: $\Pr(c_{\boldsymbol{V}})$, $\Pr(c_V)$, $\Pr(\boldsymbol{v})$, $\Pr(v)$, $\Pr(v \mid c_{\boldsymbol{E}})$
  - distributions: $\Pr(\boldsymbol{V})$, $\Pr(V)$, $\Pr(\boldsymbol{V} \mid \boldsymbol{e})$
  - distribution *sets*: $\Pr(\boldsymbol{V} \mid \boldsymbol{E})$, $\Pr(V \mid \boldsymbol{E})$

Independence of variables

Definition: Let $\boldsymbol{V}$ be a set of random variables and let $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}$. Let $\Pr$ be a joint distribution on $\boldsymbol{V}$.

The set of variables $\boldsymbol{X}$ is called conditionally independent of the set $\boldsymbol{Y}$ given the set $\boldsymbol{Z}$ in $\Pr$, if we have that

$$\Pr(\boldsymbol{X} \mid \boldsymbol{Y} \wedge \boldsymbol{Z}) = \Pr(\boldsymbol{X} \mid \boldsymbol{Z})$$

Remarks:

- the expression $\Pr(\boldsymbol{X} \mid \boldsymbol{Y} \wedge \boldsymbol{Z}) = \Pr(\boldsymbol{X} \mid \boldsymbol{Z})$ represents that $\Pr(c_{\boldsymbol{X}} \mid c_{\boldsymbol{Y}} \wedge c_{\boldsymbol{Z}}) = \Pr(c_{\boldsymbol{X}} \mid c_{\boldsymbol{Z}})$ holds for all configurations $c_{\boldsymbol{X}}$, $c_{\boldsymbol{Y}}$ and $c_{\boldsymbol{Z}}$ of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$;
- $\Pr(\boldsymbol{X} \mid \boldsymbol{Y} \wedge \boldsymbol{Z}) = \Pr(\boldsymbol{X} \mid \boldsymbol{Z}) \Rightarrow$
  $\Pr(\boldsymbol{X} \wedge \boldsymbol{Y} \mid \boldsymbol{Z}) = \Pr(\boldsymbol{X} \mid \boldsymbol{Z}) \cdot \Pr(\boldsymbol{Y} \mid \boldsymbol{Z})$ (what about $\Leftarrow$?).■

Chapter 3:

Independences and Graphical Representations

A qualitative notion of independence

Observation:

People are capable of making statements about independences among variables without having to perform numerical calculations.

Conclusion:

In human reasoning behaviour, the qualitative notion of independence is more fundamental than the quantitative notion of independence.

The (probabilistic) independence relation of a joint distribution

Definition: Let V be a set of random variables and let $\Pr$ be a joint probability distribution on V.

The independence relation $I_{\Pr}$ of $\Pr$ is a set $I_{\Pr} \subseteq \mathcal{P}(\boldsymbol{V}) \times \mathcal{P}(\boldsymbol{V}) \times \mathcal{P}(\boldsymbol{V})$, defined for all $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}$ by

$$(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}) \in I_{\Pr} \text{ if and only if } \Pr(\boldsymbol{X} \mid \boldsymbol{Y} \wedge \boldsymbol{Z}) = \Pr(\boldsymbol{X} \mid \boldsymbol{Z})$$

Remarks:

- $(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}) \in I_{\Pr}$ will be written as $I_{\Pr}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$;
  $(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}) \notin I_{\Pr}$ will be written as $\neg I_{\Pr}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$;
- a statement $I_{\Pr}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$ is called an independence statement for the joint distribution $\Pr$.

Properties of I_{Pr}: symmetry

Lemma: $I_{\mathrm{Pr}}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$ if and only if $I_{\mathrm{Pr}}(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X})$

Proof:

$$
\begin{aligned}
I_{\mathrm{Pr}}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}) &\iff \mathrm{Pr}(\boldsymbol{X} \mid \boldsymbol{Y} \wedge \boldsymbol{Z}) = \mathrm{Pr}(\boldsymbol{X} \mid \boldsymbol{Z}) \\
&\iff \frac{\mathrm{Pr}(\boldsymbol{X} \wedge \boldsymbol{Y} \wedge \boldsymbol{Z})}{\mathrm{Pr}(\boldsymbol{Y} \wedge \boldsymbol{Z})} = \frac{\mathrm{Pr}(\boldsymbol{X} \wedge \boldsymbol{Z})}{\mathrm{Pr}(\boldsymbol{Z})} \\
&\iff \frac{\mathrm{Pr}(\boldsymbol{X} \wedge \boldsymbol{Y} \wedge \boldsymbol{Z})}{\mathrm{Pr}(\boldsymbol{X} \wedge \boldsymbol{Z})} = \frac{\mathrm{Pr}(\boldsymbol{Y} \wedge \boldsymbol{Z})}{\mathrm{Pr}(\boldsymbol{Z})} \\
&\iff \mathrm{Pr}(\boldsymbol{Y} \mid \boldsymbol{X} \wedge \boldsymbol{Z}) = \mathrm{Pr}(\boldsymbol{Y} \mid \boldsymbol{Z}) \\
&\iff I_{\mathrm{Pr}}(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X})
\end{aligned}
$$

■

Properties of $I_{\Pr}$: decomposition

Lemma: $I_{\Pr}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y} \cup \boldsymbol{W}) \Rightarrow I_{\Pr}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}) \;\wedge\; I_{\Pr}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$

Proof: (sketch) (Note: $c_{\boldsymbol{Y} \cup \boldsymbol{W}} = c_{\boldsymbol{Y}} \wedge c_{\boldsymbol{W}}$!) Suppose that $\Pr(\boldsymbol{X} \mid \boldsymbol{Y} \wedge \boldsymbol{W} \wedge \boldsymbol{Z}) = \Pr(\boldsymbol{X} \mid \boldsymbol{Z})$. Then, by definition,

$$\Pr(\boldsymbol{X} \wedge \boldsymbol{Y} \wedge \boldsymbol{W} \wedge \boldsymbol{Z}) = \Pr(\boldsymbol{Y} \wedge \boldsymbol{W} \wedge \boldsymbol{Z}) \cdot \frac{\Pr(\boldsymbol{X} \wedge \boldsymbol{Z})}{\Pr(\boldsymbol{Z})}$$

For $\Pr(\boldsymbol{X} \mid \boldsymbol{Y} \wedge \boldsymbol{Z})$ we find that

$$\begin{aligned}\Pr(\boldsymbol{X} \mid \boldsymbol{Y} \wedge \boldsymbol{Z}) &= \frac{\Pr(\boldsymbol{X} \wedge \boldsymbol{Y} \wedge \boldsymbol{Z})}{\Pr(\boldsymbol{Y} \wedge \boldsymbol{Z})} \\ &= \frac{\sum_{c_{\boldsymbol{W}}} \Pr(\boldsymbol{X} \wedge \boldsymbol{Y} \wedge \boldsymbol{Z} \wedge c_{\boldsymbol{W}})}{\Pr(\boldsymbol{Y} \wedge \boldsymbol{Z})} \\ &= \frac{\Pr(\boldsymbol{X} \wedge \boldsymbol{Z})}{\Pr(\boldsymbol{Z})} = \Pr(\boldsymbol{X} \mid \boldsymbol{Z}) \qquad \blacksquare\end{aligned}$$

Properties of I_{Pr}: weak union, contraction

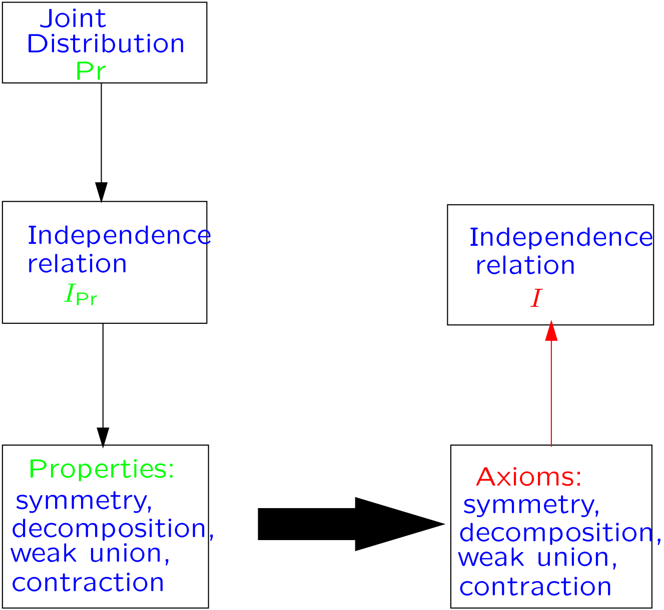
Lemma:

- if $I_{\mathrm{Pr}}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y} \cup \boldsymbol{W})$ then $I_{\mathrm{Pr}}(\boldsymbol{X}, \boldsymbol{Z} \cup \boldsymbol{W}, \boldsymbol{Y})$ (weak union);
- if $I_{\mathrm{Pr}}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$ and $I_{\mathrm{Pr}}(\boldsymbol{X}, \boldsymbol{Z} \cup \boldsymbol{W}, \boldsymbol{Y})$ then $I_{\mathrm{Pr}}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y} \cup \boldsymbol{W})$ (contraction)
- (for strictly positive Pr also the intersection property holds; see syllabus)

Proof: left as exercise 3.1.

What about $\Leftarrow$?

The definition of the independence relation



The (qualitative) independence relation I

Definition:

Let V be a set of random variables and let $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{W} \subseteq \boldsymbol{V}$.

An independence relation I on $\boldsymbol{V}$ is a ternary relation $I \subseteq \mathcal{P}(\boldsymbol{V}) \times \mathcal{P}(\boldsymbol{V}) \times \mathcal{P}(\boldsymbol{V})$ that satisfies the following properties:

- if $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$ then $I(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X})$;
- if $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y} \cup \boldsymbol{W})$ then $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$ and $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$;
- if $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y} \cup \boldsymbol{W})$ then $I(\boldsymbol{X}, \boldsymbol{Z} \cup \boldsymbol{W}, \boldsymbol{Y})$;
- if $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$ and $I(\boldsymbol{X}, \boldsymbol{Z} \cup \boldsymbol{W}, \boldsymbol{Y})$ then $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y} \cup \boldsymbol{W})$.

The first property is called the symmetry *axiom*; the second is called the decomposition axiom; the third is referred to as the weak union axiom; the last one is called contraction.

An example

Lemma:
Let I be an independence relation on a set of random variables V. We have that

$$\text{if } I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}) \text{ and } I(\boldsymbol{X} \cup \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{W}) \text{ then } I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$$

for all $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{W} \subseteq \boldsymbol{V}$.

Proof:
We observe that

$$I(\boldsymbol{X} \cup \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{W}) \Rightarrow_{\text{symm}} I(\boldsymbol{W}, \boldsymbol{Y}, \boldsymbol{X} \cup \boldsymbol{Z}) \Rightarrow_{\text{weakunion}}$$
$$\Rightarrow I(\boldsymbol{W}, \boldsymbol{Y} \cup \boldsymbol{Z}, \boldsymbol{X}) \Rightarrow_{\text{symm}} I(\boldsymbol{X}, \boldsymbol{Y} \cup \boldsymbol{Z}, \boldsymbol{W})$$

From $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$, $I(\boldsymbol{X}, \boldsymbol{Y} \cup \boldsymbol{Z}, \boldsymbol{W})$ and the contraction axiom we have that $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W} \cup \boldsymbol{Y})$; decomposition now gives $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$. ■

Representing independences

Different ways exist of representing an independence relation:

- all independence statements of the relation are explicitly stated;
- only the independence statements of a suitable subset of the relation are explicitly stated — all other statements are implicitly represented by means of the axioms;
- the independence relation is coded in a graph;
- ...

An example

Consider $\boldsymbol{V} = \{V_1, V_2, V_3, V_4\}$ and independence relation I on $\boldsymbol{V}$:

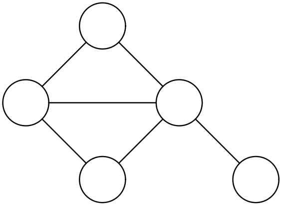
$I(\{V_1\}, \emptyset, \{V_4\})$
$I(\{V_2\}, \emptyset, \{V_4\})$
$I(\{V_3\}, \emptyset, \{V_4\})$
$I(\{V_4\}, \emptyset, \{V_1\})$
$I(\{V_4\}, \emptyset, \{V_2\})$
$I(\{V_4\}, \emptyset, \{V_3\})$
$I(\{V_1, V_2\}, \emptyset, \{V_4\})$
$I(\{V_1, V_3\}, \emptyset, \{V_4\})$
$I(\{V_2, V_3\}, \emptyset, \{V_4\})$
$I(\{V_4\}, \emptyset, \{V_1, V_2\})$
$I(\{V_4\}, \emptyset, \{V_1, V_3\})$
$I(\{V_4\}, \emptyset, \{V_2, V_3\})$
$I(\{V_1, V_2, V_3\}, \emptyset, \{V_4\})$
$I(\{V_4\}, \emptyset, \{V_1, V_2, V_3\})$
$I(\{V_1\}, \emptyset, \{V_2\})$

$I(\{V_2\}, \emptyset, \{V_1\})$
$I(\{V_1, V_4\}, \emptyset, \{V_2\})$
$I(\{V_2, V_4\}, \emptyset, \{V_1\})$
$I(\{V_2\}, \emptyset, \{V_1, V_4\})$
$I(\{V_1\}, \emptyset, \{V_2, V_4\})$
$I(\{V_2\}, \{V_1\}, \{V_4\})$
$I(\{V_3\}, \{V_1\}, \{V_4\})$
$I(\{V_2, V_3\}, \{V_1\}, \{V_4\})$
$I(\{V_4\}, \{V_1, V_2\}, \{V_3\})$
$I(\{V_2\}, \{V_1, V_3\}, \{V_4\})$
$I(\{V_4\}, \{V_1, V_3\}, \{V_2\})$
$I(\{V_1\}, \{V_2, V_3\}, \{V_4\})$
$I(\{V_4\}, \{V_2, V_3\}, \{V_1\})$
$I(\{V_4\}, \{V_3\}, \{V_1, V_2\})$
$I(\{V_4\}, \{V_3\}, \{V_1\})$

$I(\{V_4\}, \{V_1\}, \{V_2\})$
$I(\{V_4\}, \{V_1\}, \{V_3\})$
$I(\{V_4\}, \{V_1\}, \{V_2, V_3\})$
$I(\{V_1\}, \{V_2\}, \{V_4\})$
$I(\{V_3\}, \{V_2\}, \{V_4\})$
$I(\{V_1, V_3\}, \{V_2\}, \{V_4\})$
$I(\{V_4\}, \{V_2\}, \{V_1\})$
$I(\{V_4\}, \{V_2\}, \{V_3\})$
$I(\{V_4\}, \{V_2\}, \{V_1, V_3\})$
$I(\{V_1\}, \{V_3\}, \{V_4\})$
$I(\{V_2\}, \{V_3\}, \{V_4\})$
$I(\{V_1, V_2\}, \{V_3\}, \{V_4\})$
$I(\{V_1\}, \{V_4\}, \{V_2\})$
$I(\{V_2\}, \{V_4\}, \{V_1\})$
$I(\{V_3\}, \{V_1, V_2\}, \{V_4\})$

The representation of an independence relation in an undirected graph

Consider an independence relation I and an undirected graph:



the global idea is:

- represent each variable V_i by a node V_i in the graph, and v.v.;
- code the independence statements of I by means of missing edges.

The separation criterion: introduction

Definition:
Let $G = (\boldsymbol{V}_G, \boldsymbol{E}_G)$ be an undirected graph with edges $\boldsymbol{E}_G$ and nodes $\boldsymbol{V}_G = \{V_1, \ldots, V_n\}$, $n > 1$.

Let s be a path in G from a node V_i to a node V_j.

The path s is blocked by a set of nodes $\boldsymbol{Z} \subseteq \boldsymbol{V}_G$, if at least one node from $\boldsymbol{Z}$ is *on* the path s.

If s is not blocked by $\boldsymbol{Z}$, the path is called active given $\boldsymbol{Z}$.

The separation criterion

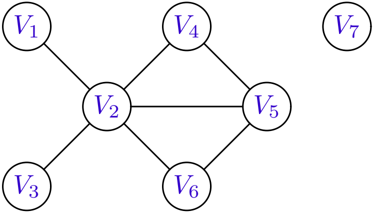
Definition:
Let $G = (\boldsymbol{V}_G, \boldsymbol{E}_G)$ be an undirected graph. Let $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}_G$ be sets of nodes in G.

The set $\boldsymbol{Z}$ separates the set $\boldsymbol{X}$ from $\boldsymbol{Y}$ in G— Notation: $\langle \boldsymbol{X} \mid \boldsymbol{Z} \mid \boldsymbol{Y} \rangle_G$— if *every simple* path in G from a node in $\boldsymbol{X}$ to a node in $\boldsymbol{Y}$ is *blocked* by $\boldsymbol{Z}$.

Remarks:

- the above notion is known as the separation criterion for undirected graphs;
- if there is no path between the nodes $\boldsymbol{X}$ and $\boldsymbol{Y}$ in a graph G, then $\langle \boldsymbol{X} \mid \emptyset \mid \boldsymbol{Y} \rangle_G$.

An example



Which of the following separation statements are valid?

a) $\langle \{V_1\} \mid \{V_2\} \mid \{V_3, V_6\} \rangle_G$

b) $\langle \{V_4\} \mid \{V_2, V_5\} \mid \{V_6\} \rangle_G$

c) $\langle \{V_4\} \mid \{V_1, V_2, V_5\} \mid \{V_6\} \rangle_G$

d) $\langle \{V_1\} \mid \{V_4\} \mid \{V_5\} \rangle_G$

e) $\langle \{V_1, V_5, V_6\} \mid \emptyset \mid \{V_7\} \rangle_G$

f) $\langle \{V_2\} \mid \{V_5\} \mid \{V_7\} \rangle_G$

g) $\langle \{V_1\} \mid \{V_5\} \mid \{V_2\} \rangle_G$

Independence relations and undirected graphs

Definition: Let I be an independence relation on a set of random variables $\boldsymbol{V}$. Let $G = (\boldsymbol{V}_G, \boldsymbol{E}_G)$ be an undirected graph with $\boldsymbol{V}_G = \boldsymbol{V}$.

- graph G is called a dependency map (D-map) for I if for all $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}$ we have:

  if $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$ then $\langle \boldsymbol{X} \mid \boldsymbol{Z} \mid \boldsymbol{Y} \rangle_G$;

- graph G is called an independency map (I-map) for I if for all $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}$ we have:

  if $\langle \boldsymbol{X} \mid \boldsymbol{Z} \mid \boldsymbol{Y} \rangle_G$ then $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$;

- graph G is called a perfect map (P-map) for I if G is both a dependency map and an independency map for I.

undirected D-maps: what can they tell?

Let I be an independence relation and G an undirected graph.

Consider a D-map for I, then

V_1 and V_2 neighbours $\Longrightarrow$ V_1, V_2 dependent

$\neg\langle\{V_1\} \mid \boldsymbol{Z} \mid \{V_2\}\rangle_G$ $\quad\quad$ $\neg I(\{V_1\}, \boldsymbol{Z}, \{V_2\})$

V_1 and V_2 non-neighbours $\Longrightarrow$??

$\langle\{V_1\} \mid \boldsymbol{Z} \mid \{V_2\}\rangle_G$

dependent

independent

conditionally independent

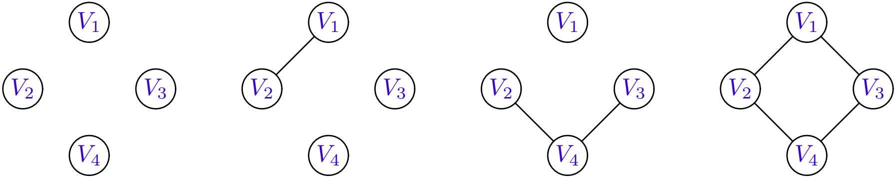
Note: statements hold for all $\boldsymbol{Z} \subseteq \boldsymbol{V}_G \setminus (\{V_1\} \cup \{V_2\})$!

An example

Consider the independence relation I on $\boldsymbol{V} = \{V_1, \ldots, V_4\}$, defined by

$$I(\{V_1\}, \{V_2, V_3\}, \{V_4\}) \text{ and } I(\{V_2\}, \{V_1, V_4\}, \{V_3\})$$

Which of the following undirected graphs are examples of D-maps for I ?



Undirected I-maps: what can they tell?

Let I be an independence relation and G an undirected graph.

Consider an I-map for I, then

V_1 and V_2 non-neighbours $\Longrightarrow$ V_1, V_2 (cond.) independent

$\langle\{V_1\} \mid \boldsymbol{Z} \mid \{V_2\}\rangle_G$ $\qquad$ $I(\{V_1\}, \boldsymbol{Z}, \{V_2\})$

V_1 and V_2 neighbours $\Longrightarrow$??

$\neg\langle\{V_1\} \mid \boldsymbol{Z} \mid \{V_2\}\rangle_G$ $\qquad$ *dependent*

independent

conditionally independent

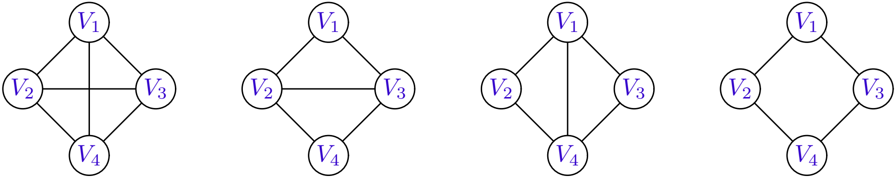
Note: statements hold for all $\boldsymbol{Z} \subseteq \boldsymbol{V}_G \setminus (\{V_1\} \cup \{V_2\})$!

An example

Consider the independence relation I on $\boldsymbol{V} = \{V_1, \ldots, V_4\}$, defined by

$$I(\{V_1\}, \{V_2, V_3\}, \{V_4\}) \text{ and } I(\{V_2\}, \{V_1, V_4\}, \{V_3\})$$

Which of the following undirected graphs are examples of I-maps for I ?



Properties of I

Let I be an independence relation on a set of random variables $\boldsymbol{V}$.

Lemma:
Every independence relation I has an undirected D-map.

Proof:
The undirected graph $G = (\boldsymbol{V}, \emptyset)$ is a D-map for I. ■

Lemma:
Every independence relation I has an undirected I-map.

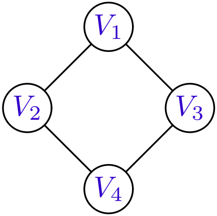
Proof:
The undirected graph $G' = (\boldsymbol{V}, \boldsymbol{V} \times \boldsymbol{V})$ is an I-map for I. ■

An example

Consider the independence relation I on $\boldsymbol{V} = \{V_1, \ldots, V_4\}$, defined by

$$I(\{V_1\}, \{V_2, V_3\}, \{V_4\}) \text{ and } I(\{V_2\}, \{V_1, V_4\}, \{V_3\})$$

The following undirected graph is a perfect map for I:



Is this P-map for I unique ?

Does every I have a P-map ?

An example

Consider an experiment with two coins and a bell: the bell sounds iff the two coins have the same outcome after a toss.

Consider: variable C_1: the outcome of tossing coin one;
variable C_2: the outcome of tossing coin two;
variable B: whether or not the bell sounds;
independence relation I for this experiment.

We have, among others, that

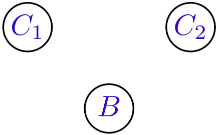
$$\begin{array}{ll} I(\{C_1\}, \emptyset, \{C_2\}) & \neg I(\{C_1\}, \{B\}, \{C_2\}) \\ I(\{C_1\}, \emptyset, \{B\}) & \neg I(\{C_1\}, \{C_2\}, \{B\}) \\ I(\{C_2\}, \emptyset, \{B\}) & \neg I(\{C_2\}, \{C_1\}, \{B\}) \end{array}$$

This independence relation is an example of an independence relation with an induced dependency.

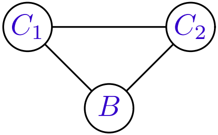
An example

Reconsider the experiment with the two coins and the bell.

- the following graph is a D-map for the independence relation I of this experiment:



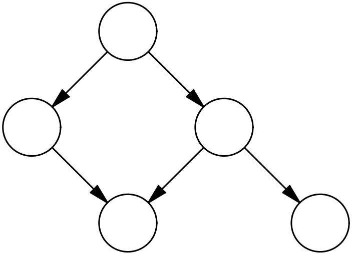
- the following graph is an I-map for I:



- Does I have a perfect map ?

The representation of an independence relation in a directed graph

Consider an independence relation I and a directed graph G:

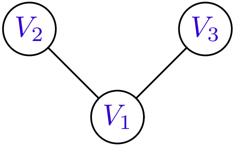


The global idea is:

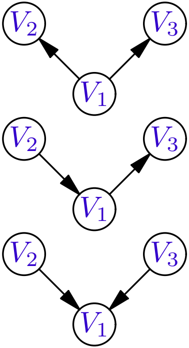
- represent each variable V_i of I by a node V_i in G, and v.v.;
- code the independence statements of I by means of missing arcs in the graph;
- use the direction of the arcs to represent induced dependencies.

Introduction

The formalism of the directed graph is more expressive than the formalism of the undirected graph:

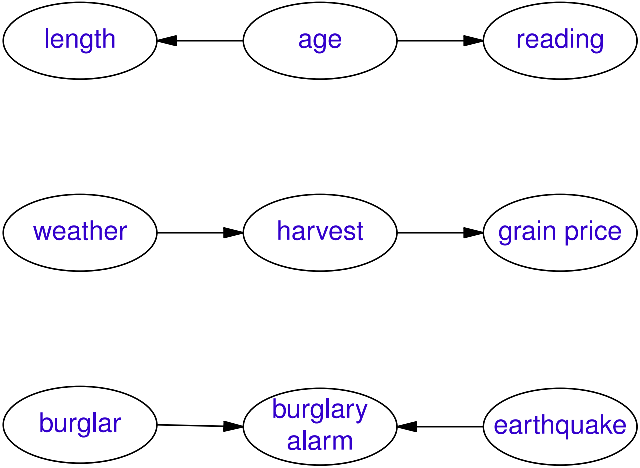


vs.



Causality ?

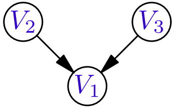
Consider the following examples:



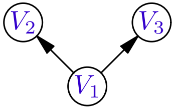
Introduction, continued

We aim to represent the following (in)dependences with directed graphs:

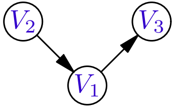
- $I(\{V_2\}, \emptyset, \{V_3\})$ and $\neg I(\{V_2\}, \{V_1\}, \{V_3\})$:



- $I(\{V_2\}, \{V_1\}, \{V_3\})$ and $\neg I(\{V_2\}, \emptyset, \{V_3\})$:



- $I(\{V_2\}, \{V_1\}, \{V_3\})$ and $\neg I(\{V_2\}, \emptyset, \{V_3\})$:

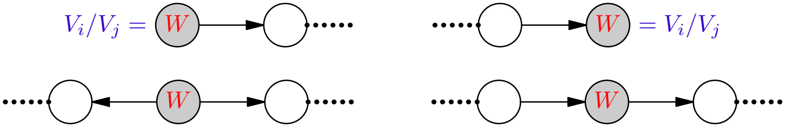


The d-separation criterion: introduction

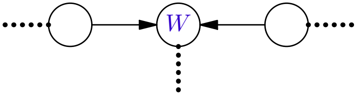
Definition: Let $G = (\boldsymbol{V}_G, \boldsymbol{A}_G)$ be an acyclic directed graph (DAG), and let s be a chain in G between V_i and $V_j \in \boldsymbol{V}_G$.

Chain s is blocked (or: in-active) by a set $\boldsymbol{Z} \subseteq \boldsymbol{V}_G$ if s contains a node W for which one of the following holds:

- $W \in \boldsymbol{Z}$ and W has at most one incoming arc *on* chain s:

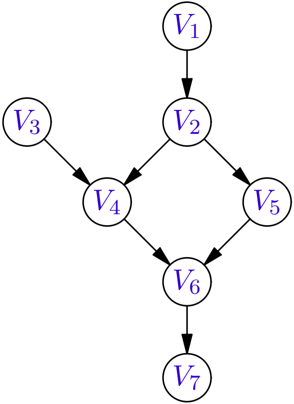


- $\boldsymbol{\sigma}^*(W) \cap \boldsymbol{Z} = \emptyset$ and W has two incoming arcs *on* chain s:



An example

Consider the following DAG and some of its chains:



1) V_4, V_2, V_5 from V_4 to V_5
2) V_1, V_2, V_5, V_6, V_7 from V_1 to V_7
3) V_3, V_4, V_6, V_5 from V_3 to V_5
4) V_2, V_4 from V_2 to V_4

Which of these chains is blocked by which of the following sets?

$$\{V_2\}, \{V_5\}, \{V_2, V_5\}, \{V_4\}, \{V_6\}, \{V_4, V_6\}$$

The d-separation criterion

Definition:

Let $G = (V_G, A_G)$ be an acyclic directed graph. Let $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}_G$ be sets of nodes in G.

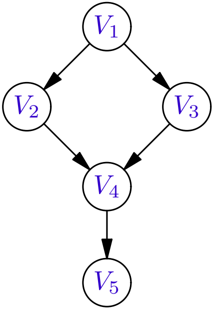
The set $\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$ in G—notation: $\langle \boldsymbol{X} \mid \boldsymbol{Z} \mid \boldsymbol{Y} \rangle_G^d$ —if *every simple chain* in G from a node in $\boldsymbol{X}$ to a node in $\boldsymbol{Y}$ is blocked by $\boldsymbol{Z}$.

Remarks:

- The above notion is known as the d-separation criterion;
- $\langle \boldsymbol{X} \mid \emptyset \mid \boldsymbol{Y} \rangle_G^d$ indicates that all chains between $\boldsymbol{X}$ and $\boldsymbol{Y}$, if any, contain a head-to-head node;
- if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are not d-separated by $\boldsymbol{Z}$, we say that they are d-connected given $\boldsymbol{Z}$.

An example

Consider the following DAG and d-separation statements:



a) $\langle \{V_1\} \mid \{V_2, V_3\} \mid \{V_5\} \rangle_G^d$

b) $\langle \{V_1\} \mid \{V_4\} \mid \{V_5\} \rangle_G^d$

c) $\langle \{V_2\} \mid \{V_1\} \mid \{V_3\} \rangle_G^d$

d) $\langle \{V_2\} \mid \{V_1, V_5\} \mid \{V_3\} \rangle_G^d$

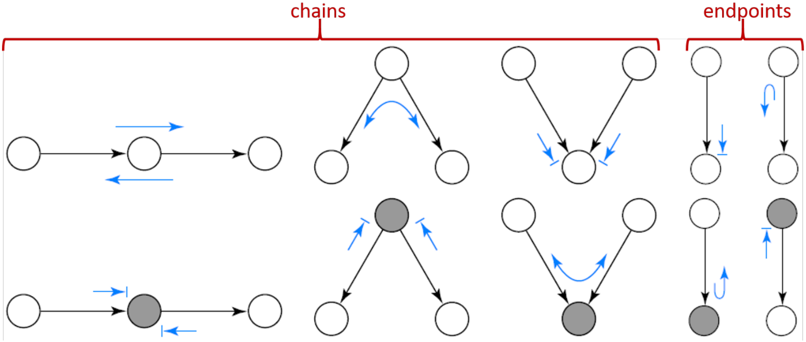
e) $\langle \{V_2\} \mid \emptyset \mid \{V_3\} \rangle_G^d$

f) $\langle \{V_1\} \mid \{V_3, V_4\} \mid \{V_2\} \rangle_G^d$

Which d-separation statements are valid in the graph ?

Bayes-Ball for determining d-separation

Determine if $\langle \boldsymbol{X} \mid \boldsymbol{Z} \mid \boldsymbol{Y} \rangle_G^d$ by dropping bouncing balls at $\boldsymbol{X}$ and following the 10 rules of Bayes-ball:



- $\boldsymbol{Z}$ is shaded
- a chain is active until a ball travelling along it meets a stop $\rightarrowtail$
- any node visited by a Bayes ball cannot be in $\boldsymbol{Y}$

Independence relations and directed graphs

Definition:
Let I be an independence relation on a set of random variables $\boldsymbol{V}$.
Let $G = (\boldsymbol{V}_G, \boldsymbol{A}_G)$ be an acyclic directed graph with $\boldsymbol{V}_G = \boldsymbol{V}$.

- the graph G is called a (directed) dependency map (D-map) for I if for every $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}$ we have that:

  if $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$ then $\langle \boldsymbol{X} \mid \boldsymbol{Z} \mid \boldsymbol{Y} \rangle_G^d$;

- the graph G is called a (directed) independency map (I-map) for I if for every $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}$ we have that:

  if $\langle \boldsymbol{X} \mid \boldsymbol{Z} \mid \boldsymbol{Y} \rangle_G^d$ then $I(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$;

- the graph G is called a (directed) perfect map (P-map) for I if G is both a dependency map and an independency map for I.

Directed D-maps: what can they tell?

Let I be an independence relation and G a DAG.

Consider a D-map for I, then

V_1 and V_2 neighbours $\Longrightarrow$ V_1, V_2 dependent

$\neg\langle\{V_1\} \mid \boldsymbol{Z} \mid \{V_2\}\rangle_G$ $\quad$ $\neg I(\{V_1\}, \boldsymbol{Z}, \{V_2\})$

V_1 and V_2 non-neighbours $\Longrightarrow$??

$\langle\{V_1\} \mid \boldsymbol{Z} \mid \{V_2\}\rangle_G$

$dependent$

$independent$

$conditionally\ dependent$ $(\boldsymbol{Z} = \emptyset)$

$conditionally\ independent$ $(\boldsymbol{Z} \neq \emptyset)$

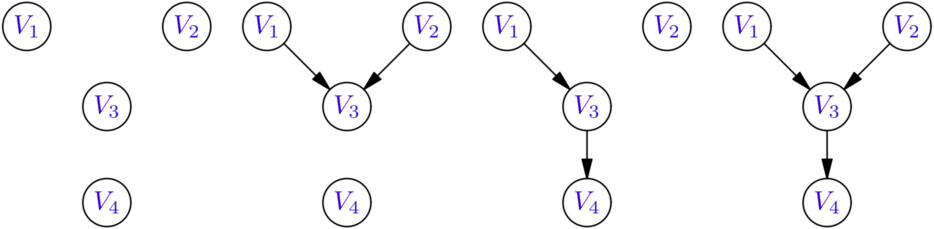
Note: statements hold for all $\boldsymbol{Z} \subseteq \boldsymbol{V}_G \setminus (\{V_1\} \cup \{V_2\})$!

An example

Consider the independence relation I on $\boldsymbol{V} = \{V_1, \ldots, V_4\}$ defined by

$$I(\{V_1\}, \emptyset, \{V_2\}) \text{ and } I(\{V_1, V_2\}, \{V_3\}, \{V_4\})$$

Which of the following DAGs are D-maps for I ?



Directed I-maps

Let I be an independence relation and G a DAG.

Consider an I-map for I, then

V_1 and V_2 non-neighbours $\Longrightarrow$ V_1, V_2 (cond.) independent, or cond. dependent (= induced)

$\langle\{V_1\} \mid \boldsymbol{Z} \mid \{V_2\}\rangle_G$ $\qquad$ $I(\{V_1\}, \boldsymbol{Z}, \{V_2\})$

V_1 and V_2 neighbours $\Longrightarrow$??

$\neg\langle\{V_1\} \mid \boldsymbol{Z} \mid \{V_2\}\rangle_G$

dependent

independent

conditionally dependent

conditionally independent

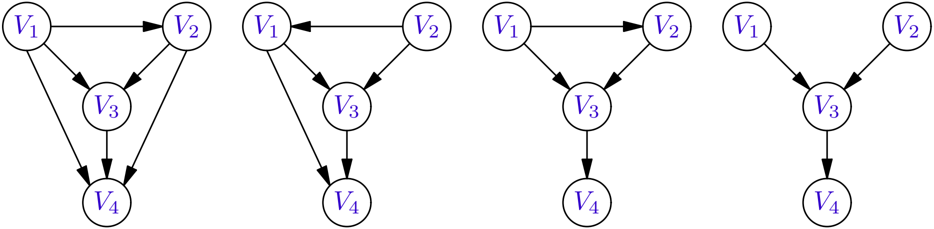
Note: statements hold for all $\boldsymbol{Z} \subseteq \boldsymbol{V}_G \setminus (\{V_1\} \cup \{V_2\})$!

An example

Consider the independence relation I on $\boldsymbol{V} = \{V_1, \ldots, V_4\}$ defined by

$$I(\{V_1\}, \emptyset, \{V_2\}) \text{ and } I(\{V_1, V_2\}, \{V_3\}, \{V_4\})$$

Which of the following DAGs are I-maps for I ?

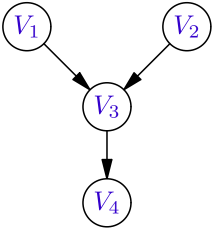


An example

Consider the independence relation I on $\boldsymbol{V} = \{V_1, \ldots, V_4\}$ defined by

$$I(\{V_1\}, \emptyset, \{V_2\}) \text{ and } I(\{V_1, V_2\}, \{V_3\}, \{V_4\})$$

The following DAG is a perfect map for I:



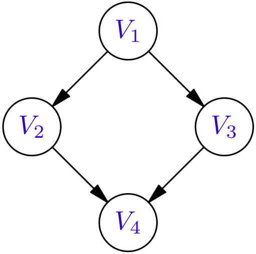
Is this P-map for I unique ?

An example

Consider the independence relation I on $\boldsymbol{V} = \{V_1, \ldots, V_4\}$ defined by

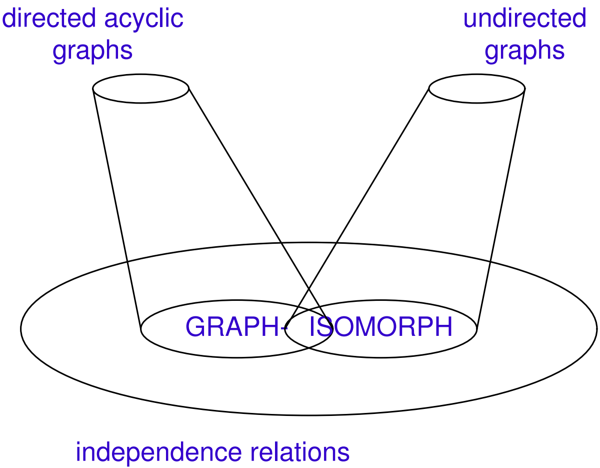
$$I(\{V_1\}, \{V_2, V_3\}, \{V_4\}) \text{ and } I(\{V_2\}, \{V_1, V_4\}, \{V_3\})$$

The relation I does not have a directed perfect map. Consider for example the following DAG G:



In graph G we have that $\langle\{V_1\} \mid \{V_2, V_3\} \mid \{V_4\}\rangle_G^d$, but also that $\langle\{V_2\} \mid \{V_1\} \mid \{V_3\}\rangle_G^d$!

Independence relations and their graphical representation



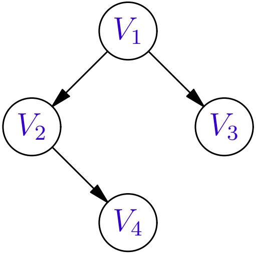
(Graph-isomorph: independence relation with perfect map.)

An I-map or a D-map ?

Reconsider the independence relation I on $\boldsymbol{V} = \{V_1, \ldots, V_4\}$ defined by

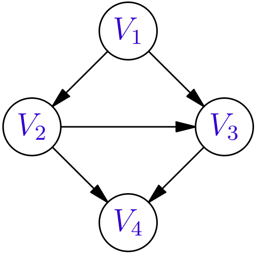
$$I(\{V_1\}, \{V_2, V_3\}, \{V_4\}) \text{ and } I(\{V_2\}, \{V_1, V_4\}, \{V_3\})$$

Compare the following two representations of independence relation I:



a D-map

and



an I-map

Recall what we were looking for...

- Compact representation of independence relation of $\Pr$;
- Factorise joint more efficiently than with chain rule $\rightarrow$ store (conditional) distributions involving less variables:

$$\begin{aligned}
\Pr(\boldsymbol{V}) &= \Pr(V_n \mid V_{n-1} \wedge \ldots \wedge V_1) \cdot \ldots \cdot \Pr(V_2 \mid V_1) \cdot \Pr(V_1) \\
&\quad \text{(chain rule)} \\
&= \ldots \\
&= \ldots \\
&= \Pr(V_n) \cdot \ldots \cdot \Pr(V_2) \cdot \Pr(V_1) \\
&\quad \text{(assuming mutual independence among all } V_i\text{)}
\end{aligned}$$

- $\Pr(X \wedge Y) = \Pr(X) \cdot \Pr(Y)$ is mathematically correct only if X is truly independent of Y

A minimal I-map

Definition: Let I be an independence relation on a set of random variables $\boldsymbol{V}$. Let $G = (\boldsymbol{V}_G, \boldsymbol{A}_G)$ be a graph with $\boldsymbol{V}_G = \boldsymbol{V}$.

The graph G is called a minimal I-map for I if the following conditions hold:

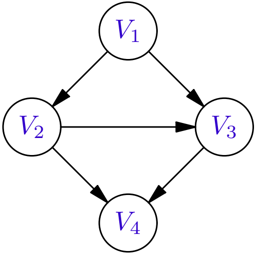
- G is an I-map for I, and
- no proper subgraph of G is an I-map for I.

An example

Consider the independence relation I on $\boldsymbol{V} = \{V_1, \ldots, V_4\}$ defined by

$$I(\{V_1\}, \{V_2, V_3\}, \{V_4\}) \text{ and } I(\{V_2\}, \{V_1, V_4\}, \{V_3\})$$

The following DAG is a minimal I-map for I:



Is this minimal I-map for I unique ?

Directed or undirected ? (I)

Directed and undirected I-maps are related.

Definition: The moral graph of a DAG $G = (\boldsymbol{V}_G, \boldsymbol{A}_G)$ is the undirected graph obtained as follows:

- for each $V_k \in \boldsymbol{V}_G$ add an edge between each pair of unconnected parents $V_i, V_j \in \boldsymbol{\rho}_G(V_k)$;
- drop the directions of all arcs.

Definition: A graph is triangulated or chordal if any loop of length ≥ 4 contains a shortcut.

Proposition: Let I be an independence relation over $\boldsymbol{V}$. Consider graphs $G = (\boldsymbol{V}_G, \boldsymbol{A}_G)$ and $G' = (\boldsymbol{V}, \boldsymbol{E}_{G'})$. Then,

$$G \text{ is an I-map for } I \quad \underset{\text{triangulation+add direction}}{\overset{\text{moralisation+drop direction}}{\Longrightarrow \atop \Longleftarrow}} \quad G' \text{ is an I-map for } I$$

Directed or undirected ? (II)

Consider independence relation $I_{\Pr}$ over $\boldsymbol{V}$ and graph G with $\boldsymbol{V} = \boldsymbol{V}_G$. Consider the following properties (partly proven later):

- Let G be a directed acyclic graph. Then G is a directed I-map of $I_{\Pr} \iff \Pr$ can be written as

$$\Pr(\boldsymbol{V}) = \prod_{V_i} \Pr(V_i \mid \boldsymbol{\rho}_G(V_i))$$

- Let G be an undirected graph. Then G is an undirected I-map of $I_{\Pr} \iff \Pr$ can be written as

$$\Pr(\boldsymbol{V}) = K \cdot \prod_{\boldsymbol{C}_i} \Phi(\boldsymbol{C}_i)$$

$\longleftarrow$ what's the meaning of these clique potentials?!?

for some normalisation factor K.