independences
and
graphical representations

Burak Galip ASLAN, PhD

probabilistic network framework

dealing with computational complexity of applying probability theory for reasoning with uncertainty in knowledge-based systems

probabilistic network framework

dealing with computational complexity of applying probability theory for reasoning with uncertainty in knowledge-based systems

the concept of conditional independence for simplifying computations

the concept of independence

the concept of independence revisited:

let $X,Y,Z \subseteq V$

the set of variables X is *conditionally independent* of the set of variables Y given the set of variables Z **if**

$$\Pr(C_{\mathbf{x}} | C_{\mathbf{y}} \wedge C_{\mathbf{z}}) = \Pr(C_{\mathbf{x}} | C_{\mathbf{z}})$$

i.e. ***once information about Z is available, information about Y is irrelevant w.r.t. X***.

otherwise

X is called *conditionally dependent* on Y given Z.

the concept of independence

with numbers; determining the independence of two sets of variables requires the computation of several probabilities and testing of several equalities

i.e. for determining independence, a joint probability distribution has to be explicitly available for the variables discerned

the concept of independence

with numbers; determining the independence of two sets of variables requires the computation of several probabilities and testing of several equalities

i.e. for determining independence, a joint probability distribution has to be explicitly available for the variables discerned

humans tend to be able to state directly, with conviction and consistency

the concept of independence

with numbers; determining the independence of two sets of variables requires the computation of several probabilities and testing of several equalities

i.e. for determining independence, a joint probability distribution has to be explicitly available for the variables discerned

humans tend to be able to state directly, with conviction and consistency

quantitative vs. qualitative

the concept of independence

to formalise properties of the qualitative concept of independence, **J. Pearl** and his co-researchers have designed **an axiomatic system for independence**

Pearl's axiomatic system for independence

independence statement:

let V be a set of statistical variables

let Pr be a joint probability distribution on V

let $\mathrm{I}_{\mathbf{Pr}}$ be the the independence relation

$\mathrm{I}_{\mathbf{Pr}} \subseteq \mathcal{P}(\mathrm{V}) \times \mathcal{P}(\mathrm{V}) \times \mathcal{P}(\mathrm{V})$ of Pr that is defined by $(\mathrm{X},\mathrm{Z},\mathrm{Y}) \in \mathrm{I}_{\mathbf{Pr}}$

if and only if $\Pr(C_{\mathbf{x}} | C_{\mathbf{y}} \wedge C_{\mathbf{z}}) = \Pr(C_{\mathbf{x}} | C_{\mathbf{z}})$ for all sets of variables $\mathrm{X},\mathrm{Y},\mathrm{Z} \subseteq \mathrm{V}$

Pearl’s axiomatic system for independence

independence statement:

let V be a set of statistical variables

let Pr be a joint probability distribution on V

let $I_{\mathbf{Pr}}$ be the the independence relation

$I_{\mathbf{Pr}}(X,Z,Y)$ denotes $(X,Z,Y) \in I_{\mathbf{Pr}}$

and

$\neg I_{\mathbf{Pr}}(X,Z,Y)$ denotes $(X,Z,Y) \notin I_{\mathbf{Pr}}$

Pearl's axiomatic system for independence

a statement $I_{\mathbf{Pr}}(X,Z,Y)$ of a joint probability distribution's independence relation $I_{\mathbf{Pr}}$ is termed as an independence statement.

in qualitative terms; an independence

info

statement $I_{\mathbf{Pr}}(X,Z,Y)$ expresses that

in the context of information about Z,

information about Y is irrelevant w.r.t. X.

Pearl's axiomatic system for independence

example statements (trivial but convenient)

$I_{\mathbf{Pr}}(X,X,Y)$ holds if $Pr(C_{\mathbf{x}}|C_{\mathbf{x}} \wedge C_{\mathbf{y}})=Pr(C_{\mathbf{x}}|C_{\mathbf{x}})$ $(1=1)$

its symmetric version also holds

$I_{\mathbf{Pr}}(Y,X,X)$ holds if $Pr(C_{\mathbf{Y}}|C_{\mathbf{x}} \wedge C_{\mathbf{x}})=Pr(C_{\mathbf{Y}}|C_{\mathbf{x}})$

Pearl’s axiomatic system for independence

Pearl builds on a set of properties that are satisfied by any joint probability distribution’s independence relation

Pearl’s axiomatic system for independence:

let V be a set of statistical variables

let Pr be a joint probability distribution on V

let $I_{\mathbf{Pr}}$ be the the independence relation

then $I_{\mathbf{Pr}}$ satisfies the following properties

Pearl's axiomatic system for independence

$$I_{\mathbf{Pr}}(X,Z,Y) \rightarrow I_{\mathbf{Pr}}(Y,Z,X)$$ (symmetry)

$$I_{\mathbf{Pr}}(X,Z,Y \cup W) \rightarrow I_{\mathbf{Pr}}(X,Z,Y) \wedge I_{\mathbf{Pr}}(X,Z,W)$$ (decomposition)

$$I_{\mathbf{Pr}}(X,Z,Y \cup W) \rightarrow I_{\mathbf{Pr}}(X,Z \cup W,Y)$$ (weak union)

$$I_{\mathbf{Pr}}(X,Z,Y) \wedge I_{\mathbf{Pr}}(X,Z \cup Y,W) \rightarrow I_{\mathbf{Pr}}(X,Z,Y \cup W)$$ (contraction)

Pearl's axiomatic system for independence

for all mutually disjoint sets $(A \cap B = \emptyset)$ of variables X,Y,Z,W⊆V, if the distribution **Pr** is strictly positive, then $I_{\mathbf{Pr}}$ also satisfies the additional property:

$$I_{\mathbf{Pr}}(X, Z \cup W, Y) \wedge I_{\mathbf{Pr}}(X, Z \cup Y, W) \rightarrow I_{\mathbf{Pr}}(X, Z, Y \cup W)$$

(intersection)

it holds for all mutally disjoint sets of variables $X,Y,Z,W \subseteq V$

Pearl’s axiomatic system for independence

Pearl’s theorem holds for mutually disjoint sets of variables only [Pearl, 1988]

However also **holds for overlapping sets of variables** [Van der Gaag & Meyer, 1998]

informational independence

let V be a set of statistical variables

A **semi-graphoid** independence relation on V is a ternary relation $\mathrm{I} \subseteq P(V) \times P(V) \times P(V)$ satisfying following properties for all sets of variables $X,Y,Z,W \subseteq V$

$\mathrm{I}(X,Z,Y) \rightarrow \mathrm{I}(Y,Z,X)$ (symmetry)

$\mathrm{I}(X,Z,Y \cup W) \rightarrow \mathrm{I}(X,Z,Y) \wedge (X,Z,W)$ (decomposition)

$\mathrm{I}(X,Z,Y \cup W) \rightarrow \mathrm{I}(X,Z \cup W,Y)$ (weak union)

$\mathrm{I}(X,Z,Y) \wedge \mathrm{I}(X,Z \cup Y,W) \rightarrow \mathrm{I}(X,Z,Y \cup W)$ (contraction)

informational independence

a **graphoid** independence relation on V is a semi-graphoid independence relation on V such that it satisfies the below additional property as well for all sets of variables $X,Y,Z,W \subseteq V$

$\mathrm{I}(X,Z \cup W,Y) \wedge \mathrm{I}(X,Z \cup Y,W) \rightarrow \mathrm{I}(X,Z,Y \cup W)$ (intersection)

«learning irrelevant information does <u>not</u> alter the independences among the variables discerned»

informational independence

qualitative meanings of axioms;

- ***symmetry axiom***
- *decomposition axiom*
- *weak union axiom*
- *contraction axiom*
- *intersection axiom*

symmetry axiom

$I(X,Z,Y) \rightarrow I(Y,Z,X)$ (for all sets of variables $X,Y,Z \subseteq V$)

if information about Y is deemed irrelevant w.r.t. X <u>in the context of some information about Z</u>, then information about X must be irrelevant w.r.t. Y <u>in this context</u>.

informational independence

qualitative meanings of axioms;

- *symmetry axiom*
- ***decomposition axiom***
- *weak union axiom*
- *contraction axiom*
- *intersection axiom*

decomposition axiom

$$I(X,Z,Y\cup W)\rightarrow I(X,Z,Y)\wedge(X,Z,W)$$

(for all sets of variables $X,Y,Z,W\subseteq V$)

if information about both Y and W is judged irrelevant w.r.t. X <u>in the context of some information about Z</u>, then both information about Y and information about W must be irrelevant w.r.t. X separately <u>in the same context</u>.

it is Pearl’s original axiom above, however a reformulation like below is also possible:

$I(X,Z,Y\cup W)\rightarrow I(X,Z,Y)$ (for all sets of variables $X,Y,Z,W\subseteq V$)

$I(X,Z,Y\cup W)\rightarrow I(X,Z,W)$ (for all sets of variables $X,Y,Z,W\subseteq V$)

informational independence

qualitative meanings of axioms;

- *symmetry axiom*
- *decomposition axiom*
- ***weak union axiom***
- *contraction axiom*
- *intersection axiom*

weak union axiom

$I(X,Z,Y \cup W) \rightarrow I(X,Z \cup W,Y)$
(for all sets of variables $X,Y,Z,W \subseteq V$)

learning some information about W that is known to be irrelevant w.r.t. X, <u>in the context of some information about Z</u>, **can<u>not</u> help irrelevant information about Y to become relevant w.r.t. X** <u>in the same context</u>.

informational independence

qualitative meanings of axioms;

- *symmetry axiom*
- *decomposition axiom*
- *weak union axiom*
- ***contraction axiom***
- *intersection axiom*

contraction axiom

$$I(X,Z,Y) \wedge I(X,Z \cup Y,W) \rightarrow I(X,Z,Y \cup W)$$

(for all sets of variables $X,Y,Z,W \subseteq V$)

if we judge information about W to be irrelevant w.r.t. X <u>after learning some irrelevant information about Y</u>, then **the information about W must have been irrelevant w.r.t. X <u>before</u> we learned something about Y**.

conditional reverse of the weak union axiom also holds as a reformulation:

$$I(X,Z,Y \cup W) \rightarrow I(X,Z,Y) \wedge I(X,Z \cup Y,W)$$

(for all sets of variables $X,Y,Z,W \subseteq V$)

informational independence

qualitative meanings of axioms;

- *symmetry axiom*
- *decomposition axiom*
- *weak union axiom*
- *contraction axiom*
- ***intersection axiom***

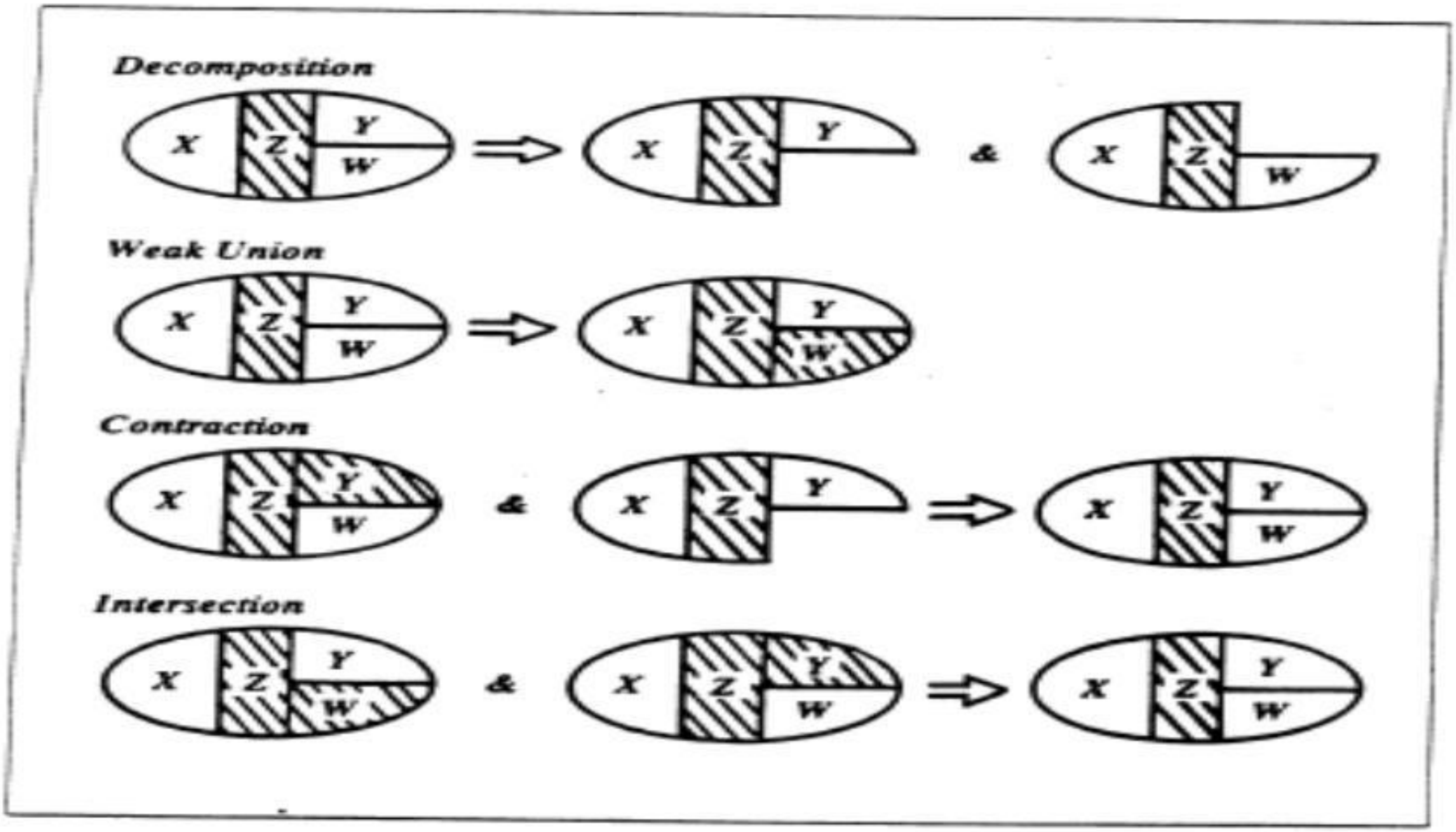
intersection axiom

$$I(X,Z\cup W,Y) \wedge I(X,Z\cup Y,W) \rightarrow I(X,Z,Y\cup W)$$

(for all sets of variables $X,Y,Z,W \subseteq V$)

if <u>in the context of some information about Z</u>, learning some information about W renders information about Y irrelevant w.r.t. X, <u>**and**</u> learning some information about Y renders W irrelevant w.r.t. X; then **information about both Y and W must be irrelevant w.r.t. X** <u>given Z</u>.

informational independence overview



Graphical interpretation of the axioms governing conditional independence.

graphical representations of independence

one of the main problems of applying probability theory for *automated reasoning with uncertainty in a knowledge-based system* is the **space complexity** of representing a joint probability distribution.

graphical representations of independence

one of the main problems of applying probability theory for *automated reasoning with uncertainty in a knowledge-based system* is the **space complexity** of representing a joint probability distribution.

since the **concept of independence** plays a key role in solving this problem, a formalism for **representing** joint probability distributions should allow for efficiently modeling independencies.

representing an independence relation

- enumerating the separate statements of an independence relation explicitly?

this is impractical as the number of tuples in an independence relation can be astronomical

representing an independence relation

• using Pearl’s axioms for only the statements from an appropriate subset of the independence relation are enumareted explicitly and all of its other statements are defined implicitly by this set and the defining axioms.

• **graphical encoding** for more concise representations of independence relations

this approach is a more economical representation for an independence relation than an explicit enumeration, however it can still require exponential space

undirected graphs

probabilistic meaning has to be assigned to the topological properties of a graph

undirected graphs

probabilistic meaning has to be assigned to the topological properties of a graph

- variables -> vertices
- independence statements -> absence of edges

undirected graphs

probabilistic meaning has to be assigned to the topological properties of a graph

- variables -> vertices
- independence statements -> absence of edges

separation criterion: blocking all paths between two given sets of vertices

undirected graphs

let $G=(V(G),E(G))$ be an undirected graph

let $X,Y,Z \subseteq V(G)$ be sets of vertices in G

the set of vertices Z is said to separate the set of vertices X and Y in G – that is denoted as $\langle X,Z,Y \rangle_G$ if for each vertex $V_i \in X$ and $V_j \in Y$ every simple path (i.e. all vertices must be distinct) from V_i to V_j in G contains **at least one vertex** from Z.

if Z separates X from Y, then Z blocks **any flow of information or influence** between X and Y i.e. X and Y are <u>conditionally independent</u> given Z.

undirected graphs

for all sets of variables $X,Y,Z \subseteq V$

- *D-map (undirected dependence map)*

  if $I(X,Z,Y)$ then $<X,Z,Y>_{\mathbf{G}}$

(hint: add edges starting from trivial D-map until no independence statement is violated)

- *I-map (undirected independence map)*

  if $<X,Z,Y>_{\mathbf{G}}$ then $I(X,Z,Y)$

(hint: remove edges starting from trivial I-map until no new independence statement is defined)

- *P-map (undirected perfect map)*

if G is both an undirected D-map and an undirected I-map for I at the same time.

undirected graphs

- *D-map (undirected dependence map)*
- *I-map (undirected independence map)*
- *P-map (undirected perfect map)*

every independence relation has an undirected D-map and undirected I-map

not every independence relation has an undirected P-map

an independence relation for which an undirected P-map exists is termed *undirected graph-isomorphic*

directed graphs

not every independence relation can be represented effectively in undirected graph formalism

directed graphs (digraphs) have a relatively stronger expressive power for representing independence relations (acyclic digraphs)

directed graphs

- *variables -> vertices*
- *independence statements -> absence of arcs*
- *d-separation criterion*

let $G=\big(V(G),A(G)\big)$ be an ayclic digraph

let **s** be a chain in G between $V_i \in V(G)$ and $V_j \in V(G)$

directed graphs

the chain **s** is blocked by a (possibly empty) set of vertices $W \subseteq V(G)$ if $V_i \in W$ or $V_j \in W$ or s contains 3 consecutive vertices for which (at least) one of the following conditions is valid:

a. arcs (X_2,X_1) and (X_2,X_3) are on the chain S; $X_2 \in W$

b. arcs (X_1,X_2) and (X_2,X_3) are on the chain S; $X_2 \in W$

c. arcs (X_1,X_2) and (X_3,X_2) are on the chain S; and $\sigma^*(X_2) \cap W = \emptyset$ *(all successors of vertex X_2 in G)*
(this is called **induced independence**)

directed graphs

the formalism of directed graphs allows for distinguishing between three alternatives since there are **three different ways** of in which *two arcs between three vertices* can be directed

where (a) and (b) are two conditions of the concept of *blocking* and (c) condition models an *induced independence*

d-separation criterion *(instead of «separation criterion» in undirected graphs)* and the concept of independence

d-separation criterion

let $G=\mathbf{(}V(G),A(G)\mathbf{)}$ be an ayclic digraph

let $X,Y,Z \subseteq V(G)$ be sets of vertices in G

The set of vertices Z is said to d-separate the sets of vertices X and Y in G, which is denoted as $\langle X|Z|Y\rangle^{\mathbf{d}}_{\mathbf{G}}$ if each vertex $V_{\mathbf{i}} \in X$ and $V_{\mathbf{j}} \in Y$; every chain from $V_{\mathbf{i}}$ to $V_{\mathbf{j}}$ in G is blocked by Z.

map definitions

D-map: for an independence relation, any pair of **neighboring vertices** represents a pair of dependent variables; however, not every pair of dependent variables needs to represented as a pair of neighboring vertices

I-map: for an independence relation, any pair of **non-neighboring** vertices represents a pair of variables that are independent; however, not every pair of independent variables is represented as a pair of non-neighboring vertices

P-map: a directed P-map is both a directed D-map and a directed I-map

directed graphs

let V be a set of statistical variables

let I be an independence relation on V

let G=**(**V(G),A(G)**)** be an acyclic digraph with V(G)=V

for all sets of variables $X,Y,Z \subseteq V$

- *D-map (directed dependence map)*

  if $I(X,Z,Y)$ then $\langle X,Z,Y\rangle_{\mathbf{G}}^{\mathbf{d}}$

(hint: add arcs starting from trivial directed D-map until no independence statement is violated)

- *I-map (directed independence map)*

  if $\langle X,Z,Y\rangle_{\mathbf{G}}^{\mathbf{d}}$ then $I(X,Z,Y)$

(hint: remove arcs starting from trivial directed I-map until no new independence statement is defined)

- *P-map (directed perfect map)*

(if G is both a directed D-map and a directed I-map for I at the same time)

directed graphs

for every independence relation there exists a directed D-map and directed I-map, however not every independence relation has a directed P-map

an independence relation **I** is said to be directed graph-isomorphic if there exists an acyclic digraph G such that G is a directed P-map for **I**

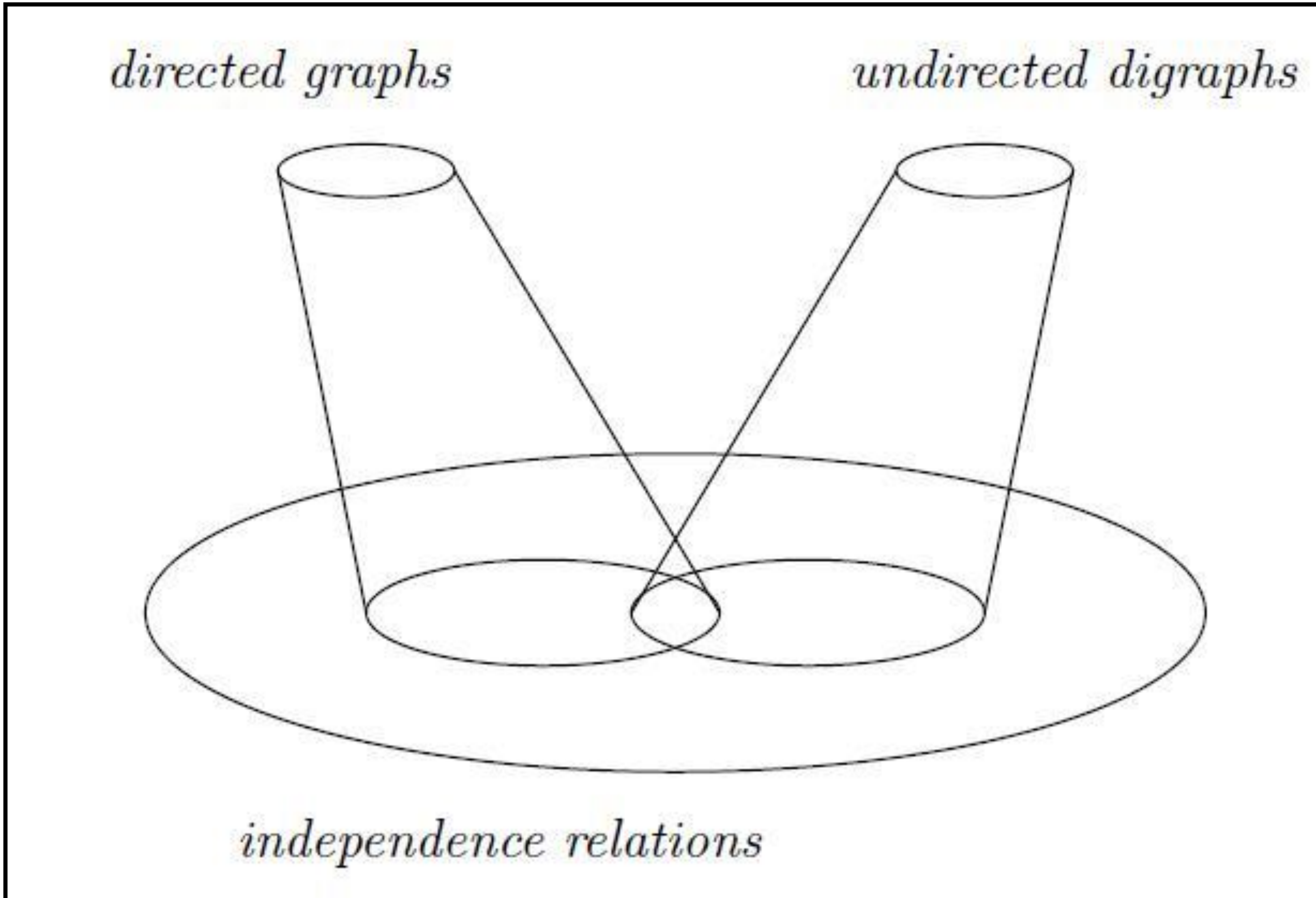
if an independence relation is directed graph-isomorphic, it may allow several different P-maps, that is, a directed P-map does not need to be unique

choosing a graphical representation

both formalisms allow for a *space-efficicient* representation of an independence relation (polynomial w.r.t. # of variables at hand)

efficiently verifying independence statements w/o requiring numerical computations

choosing a graphical representation



choosing a graphical representation

there are independence relations that are undirected graph-isomorphic yet not directed graph-isomorphic; and vice versa. *(e.g. There is an undirected P-map but no directed P-map or there is a directed P-map but no undirected P-map)*

choosing a graphical representation

there are independence relations that are undirected graph-isomorphic yet not directed graph-isomorphic; and vice versa. *(e.g. There is an undirected P-map but no directed P-map or there is a directed P-map but no undirected P-map)*

also there are independence relations that are both undirected and directed graph-isomorphic. *(i.e. the independence relation has both an undirected P-map and a directed P-map)*

choosing a graphical representation

there are independence relations that are undirected graph-isomorphic yet not directed graph-isomorphic; and vice versa. *(e.g. There is an undirected P-map but no directed P-map or there is a directed P-map but no undirected P-map)*

also there are independence relations that are both undirected and directed graph-isomorphic. *(i.e. the independence relation has both an undirected P-map and a directed P-map)*

and there are independence relations that are not graph-isomorphic at all. *(i.e. no P-map at all)*

choosing a graphical representation

this property negatively affects the suitability of graphical formalisms for representing independence relations

the efficiency of representation, however, generally is considered to outweigh the lack of expressive power of these formalism

choosing a graphical representation

this property negatively affects the suitability of graphical formalisms for representing independence relations

the efficiency of representation, however, generally is considered to outweigh the lack of expressive power of these formalism

it is possible to mix directed and undirected graphs *(chain graphs)* in order to increase the capability of representing a broader class of independence relations, however, they are hardly used for practical applications as a result of their relatively complex semantics

choosing a graphical representation

in the probabilistic network framework, therefore, a graphical formalism is used for representing independencies

the independence relations that are typically encountered in practical problem domains often are best represented by a **directed graph** rather than by an undirected graph, the formalism of directed graphs is employed in the framework

choosing a graphical representation

in the probabilistic network framework, therefore, a graphical formalism is used for representing independencies

the independence relations that are typically encountered in practical problem domains often are best represented by a **directed graph** rather than by an undirected graph, the formalism of directed graphs is employed in the framework

there does exist an undirected counterpart of the probabilistic network framework called ***Markov network***

choosing a graphical representation

in real-life problem domains, independence relations encountered may not be graph-isomorphic; for such an independence relation, **it is not possible to faithfully represent all independences as well as dependences**

the relation therefore has to be represented in either a directed I-map or directed D-map

choosing a graphical representation

in real-life problem domains, independence relations encountered may not be graph-isomorphic; for such an independence relation, **it is not possible to faithfully represent all independences as well as dependences**

the relation therefore has to be represented in either a directed I-map or directed D-map

the main idea is **exploiting independencies for simplifying computations**, so independences could be read from the graphical representation of an independence relation

choosing a graphical representation

an independence relation that is not directed graph-isomorphic therefore is best represented by a **directed I-map**

as many of the independences of the relation as possible are modeled by keeping the number of unrepresented independences at minimum

choosing a graphical representation

an independence relation that is not directed graph-isomorphic therefore is best represented by a **directed I-map**

as many of the independences of the relation as possible are modeled by keeping the number of unrepresented independences at minimum

a directed I-map that does not contain any superflous arcs is called ***minimal I-map***

minimal directed I-map

let V be a set of statistical variables

let I be an independence relation on V

let G=**(**V(G),A(G)**)** be an acyclic digraph with V(G)=V

then, the digraph G is called a **“minimal directed I-map”** for I if G is a directed I-map for I and no proper subgraph of G is a directed I-map for I.

wrap-up

this chapter formalises two types of independence relations. The first, $I_{\mathbf{Pr}}$, is the type of relation that can be captured by a probability distribution. These independence relations form a proper subset of a more general type of independence relation I that can be **abstracted** away from probability distributions.

the chapter also discusses different representations of independence relations, most notably (in)directed graphs. An important notion introduced in this chapter is the concept of **d-separation**.

wrap-up

this chapter formalises two types of independence relations. The first, $I_{\mathbf{Pr}}$, is the type of relation that can be captured by a probability distribution. These independence relations form a proper subset of a more general type of independence relation I that can be **abstracted** away from probability distributions.

the chapter also discusses different representations of independence relations, most notably (in)directed graphs. An important notion introduced in this chapter is the concept of **d-separation**.

independence relations and their representation are still an area of ongoing research.

references

Probabilistic Reasoning

(Probabilistisch Redeneren)

authors: Linda van der Gaag
Silja Renooij

Fall 2017

Universiteit Utrecht

WIKIPEDIA

Google™