Bayesian Networks

II. Probabilistic Independence and Separation in Graphs

Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)
Institute for Business Economics and Information Systems
& Institute for Computer Science
University of Hildesheim
http://www.ismll.uni-hildesheim.de

Probability spaces

Definition 1. Let Ω be a finite set. We call Ω the **sample space** and every subset $E \subseteq \Omega$ an **event**; subsets containing exactly one element, i.e.

$$E = \{e\}, \quad e \in \Omega$$

are called **elementary events**.

A function

$$p : \mathcal{P}(\Omega) \rightarrow [0, 1]$$

with

1. p is additive, i.e. for disjunct $E, F \subseteq \Omega$:
$$p(E \cup F) = p(E) + p(F)$$
2. $p(\Omega) = 1$

is called **probability function** (axioms of probability, Kolmogorov, 1933). A pair (Ω, p) is called **probability space**.

Lemma 1.

$$p(E) = \sum_{e \in E} p(\{e\}), \quad E \subseteq \Omega$$

Example 1. Throwing a dice can be described by

$$\Omega := \{1, 2, 3, 4, 5, 6\}$$

For a fair dice we have

$$p(\{1\}) = p(\{2\} = \ldots = p(\{6\}) = \frac{1}{6}$$

Then $E = \{2\}$ is the event of dicing a 2, $F = \{2, 4, 6\}$ the event of dicing an even number.

$$p(\{2, 4, 6\}) = p(\{2\}) + p(\{4\}) + p(\{6\}) = \frac{1}{2}$$

Conditional independent events (1/2)

Definition 2. Let $E, F \subseteq \Omega$ with $p(F) > 0$. Then

$$p(E|F) := \frac{p(E \cap F)}{p(F)}$$

is called **conditional probability** of E given F.

Two events $E, F \subseteq \Omega$ are called **independent**, if

$$p(E \cap F) = p(E) \cdot p(F)$$

i.e., if $p(E|F) = p(E)$ or $p(E) = 0$ or $p(F) = 0$.

Example 2. Let $F := \{2, 4, 6\}$ be the event of dicing an even number. Then the conditional probability

$$p(\{2\}|F) = \frac{1}{6} / \frac{1}{2} = \frac{1}{3}$$

describes the probability of dicing a 2 given we diced an even number.

Example 3. The events $E := \{2, 4, 6\}$ of dicing an even number and $F := \{1, 2, 3, 4\}$ of dicing a number less than 5 are independent as

$$p(E \cap F) = p(\{2, 4\}) = \frac{1}{3}$$
$$\stackrel{!}{=} p(E) \cdot p(F) = \frac{1}{2} \cdot \frac{2}{3}$$

Conditional independent events (2/2)

Definition 3. Let $G \subseteq \Omega$ be an event with $p(G) > 0$. Two events $E, F \subseteq \Omega$ are called **conditionally independent** given G, if

$$p(E \cap F \cap G) = p(E \cap G) \cdot p(F \cap G)/p(G)$$

i.e., if $p(E|F \cap G) = p(E|G)$ or $p(E|G) = 0$ or $p(F|G) = 0$.

Definition 4. A partition $(E_i)_{i=1,\ldots,m}$ of Ω is also called **a set of mutually exclusive and exhaustive events**, i.e.

1. $E_i \neq \emptyset$,
2. $\bigcup_{i=1}^m E_i = \Omega$, and
3. E_i are pairwise disjunct (i.e., $E_i \cap E_j = \emptyset$ for $i \neq j$).

Example 4. The events

- $E := \{2, 4, 6\}$ of dicing an even number and
- $F := \{1, 2, 3, 4, 5\}$ of dicing anything but 6

are dependent as

$$p(E \cap F) = p(\{2, 4\}) = \frac{1}{3} \overset{!}{\neq} p(E) \cdot p(F) = \frac{1}{2} \cdot \frac{5}{6}$$

But given the event

- $G := \{1, 2, 3, 4\}$ of dicing a number less than 5,

E and F are conditionally independent given G as

$$p(E \cap F \cap G) = p(\{2, 4\}) = \frac{1}{3} \overset{!}{=} p(E \cap G) \cdot p(F \cap G)/p(G)$$

Bayes' Theorem

Theorem 1 (Bayes, 1763). *Let $E, F \subseteq \Omega$ be two events with $p(E), p(F) > 0$. Then*

$$p(E|F) = \frac{p(F|E) \cdot p(E)}{p(F)}$$

Let $(E_i)_{i=1,\ldots,m}$ be a partition of Ω with $p(E_i) > 0$ for all i. Then

$$p(E_j|F) = \frac{p(F|E_j) \cdot p(E_j)}{\sum_{i=1}^m p(F|E_i) \cdot p(E_i)}$$

Example 5. Assign each object in fig. 1 an equal probability $\frac{1}{13}$. Let $E_1 =$ "label is one", $E_2 =$ "label is two", and $F =$ "color is black". Then

$$p(E_1|F) = \frac{p(F|E_1)p(E_1)}{p(F|E_1)p(E_1) + p(F|E_2)p(E_2)} = \frac{\frac{3}{5} \cdot \frac{5}{13}}{\frac{3}{5} \cdot \frac{5}{13} + \frac{6}{8} \cdot \frac{8}{13}} = \frac{1}{3}$$

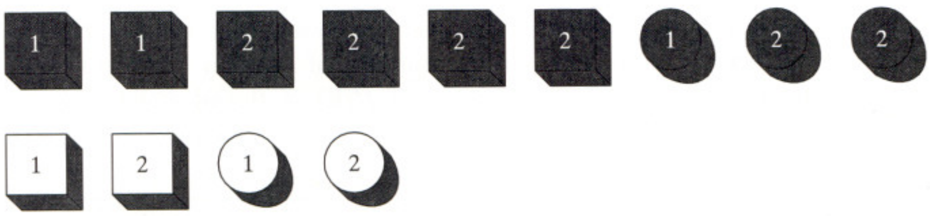


Figure 1: 13 objects with different shape, color, and label [Nea03, p. 8].

Random variables and probability distributions

Definition 5. Any function

$$X : \Omega \to X$$

is called a **random variable** (by abuse of notation we label both, the map and the target space with X).

We assign each value $x \in X$ a probability via

$$p(X = x) := p(X^{-1}(x))$$

p is called the **probability distribution of** X.

If X is numeric, e.g., $X = \mathbb{R}$, we call

$$E(X) := \sum_{x \in X} x \cdot p(x)$$

the **expected value** of X.

Example 6. Let Ω contain the outcomes of a throw of two (distinguishable) dice, i.e.

$$\Omega := \{(1,1), (1,2), \dots, (1,6), \\ (2,1), (2,2), \dots, (6,5), (6,6)\}$$

Then the sum of the two dice,

$$X : \begin{array}{rcl} \Omega & \to & \mathbb{N} \\ (i,j) & \mapsto & i+j \end{array}$$

is a random variable.

The value $X = 3$ then represents the event $X^{-1}(3) = \{(1,2),(2,1)\}$ and thus $p(X = 3) = \frac{2}{36}$.

The expected value of X is $E(X) = 7$.

X	2	3	4	5	6	7	8	9	10	11	12
$p(X)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Joint probability distributions

Definition 6. Let X and Y be two random variables. Then their cartesian product

$$X \times Y : \begin{array}{rcl} \Omega & \to & X \times Y \\ e & \mapsto & (X(e), Y(e)) \end{array}$$

is again a random variable; its distribution is called **joint probability distribution** of X and Y.

Example 7. Let Ω be the outcomes of a throw of two dices and X the sum of their numbers as before. Let Y be

$$Y(i,j) := \begin{cases} \text{odd}, & \text{if } i \text{ and } j \text{ is odd} \\ \text{even}, & \text{if } i \text{ or } j \text{ is even} \end{cases}$$

Then the probability of

$$p(X = 4, Y = \text{odd}) = p(\{(1,3),(3,1)\}) = \frac{2}{36}$$

In general,

$$p(X = x, Y = y) \neq p(X = x) \cdot p(Y = y)$$

as can be seen here:

$$p(X = 4) = p(\{(1,3),(3,1),(2,2)\}) = \frac{3}{36}$$

$$p(Y = \text{odd}) = \frac{9}{36}$$

Marginal probability distributions

Definition 7. Let p be a the joint probability of two random variables X and Y,

$$X \times Y : \Omega \rightarrow X \times Y$$

Then

$$p(X = x) := p^{\downarrow X}(x) := \sum_{y \in Y} p(X = x, Y = y)$$

is a probability distribution of X called **marginal probability distribution**.

Example 8. Assume the joint probability distribution of four random variables P (pain), W (weightloss), V (vomiting) and A (adeno) given in fig. 2.

Then the marginal distribution of V and A is

Vomiting	Y	N
Adeno Y	0.350	0.350
N	0.090	0.210

Pain	Y				N			
Weightloss	Y		N		Y		N	
Vomiting	Y	N	Y	N	Y	N	Y	N
Adeno Y	0.220	0.220	0.025	0.025	0.095	0.095	0.010	0.010
N	0.004	0.009	0.005	0.012	0.031	0.076	0.050	0.113

Figure 2: Joint probability distribution of four random variables P (pain), W (weightloss), V (vomiting) and A (adeno).

Marginal probability distributions / example

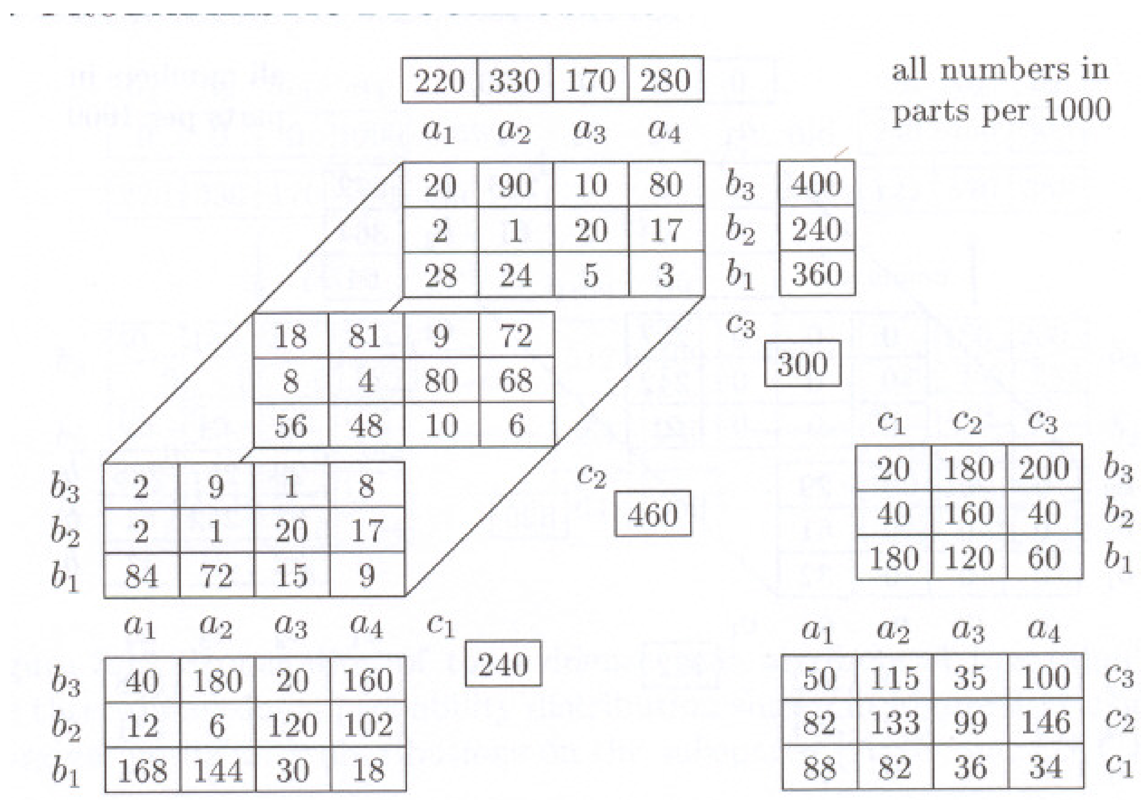


Figure 3: Joint probability distribution and all of its marginals [BK02, p. 75].

Independent variables

Definition 8. Let $\mathcal{X}, \mathcal{Y}$ be sets of variables. By abuse of notation we write $\mathcal{X} = x$ for a tuple $(x_X)_{X \in \mathcal{X}}$ of values $x_X \in X$.

$\mathcal{X}, \mathcal{Y}$ are called **independent sets of variables**, when all pairs of events $\mathcal{X} = x$ and $\mathcal{Y} = y$ are independend, i.e.

$$p(\mathcal{X} = x, \mathcal{Y} = y) = p(\mathcal{X} = x) \cdot p(\mathcal{Y} = y)$$

for all x and y or equivalently

$$p(\mathcal{X} = x | \mathcal{Y} = y) = p(\mathcal{X} = x)$$

for y with $p(\mathcal{Y} = y) > 0$.

Example 9. Let Ω be the cards in an ordinary deck and

- R be the variable that is true (Y), if a card is royal,
- T be the variable that is true (Y), if a card is a ten or a jack, and
- S be the variable that is true (Y), if a card is spade.

S	R	T	$p(R,T\|S)$
Y	Y	Y	1/13
		N	2/13
	N	Y	1/13
		N	9/13
N	Y	Y	3/39 = 1/13
		N	6/39 = 2/13
	N	Y	3/39 = 1/13
		N	27/39 = 9/13

R	T	$p(R,T)$
Y	Y	4/52 = 1/13
	N	8/52 = 2/13
N	Y	4/52 = 1/13
	N	36/52 = 9/13

Conditionally independent variables

Definition 9. Let $\mathcal{X}, \mathcal{Y}$ be sets of variables. Let $\mathcal{Z}$ be a third set of variables. $\mathcal{X}, \mathcal{Y}$ are called **conditionally independent sets of variables given** $\mathcal{Z}$, when for all events $\mathcal{Z} = z$ with $p(\mathcal{Z} = z) > 0$ all pairs of events $\mathcal{X} = x$ and $\mathcal{Y} = y$ are conditionally independend given $\mathcal{Z} = z$, i.e.

$$p(\mathcal{X} = x, \mathcal{Y} = y, \mathcal{Z} = z) = p(\mathcal{X} = x, \mathcal{Z} = z) \cdot p(\mathcal{Y} = y, \mathcal{Z} = z)/p(\mathcal{Z} = z)$$

for all x, y and z (with $p(\mathcal{Z} = z) > 0$), or equivalently

$$p(\mathcal{X} = x|\mathcal{Y} = y, \mathcal{Z} = z) = p(\mathcal{X} = x|\mathcal{Z} = z)$$

We write $I_p(\mathcal{X}, \mathcal{Y}|\mathcal{Z})$ for the statement, that $\mathcal{X}$ and $\mathcal{Y}$ are conditionally independent given $\mathcal{Z}$.

Conditionally independent variables

Example 10. Assume S (shape), C (color), and L (label) be three random variables that are distributed as shown in figure 4.

We show $I_p(\{L\}, \{S\}|\{C\})$, i.e., that label and shape are conditionally independent given the color.

C	S	L	$p(L\|C, S)$
black	square	1	2/6 = 1/3
		2	4/6 = 2/3
	round	1	1/3
		2	2/3
white	square	1	1/2
		2	1/2
	round	1	1/2
		2	1/2

C	L	$p(L\|C)$
black	1	3/9 = 1/3
	2	6/9 = 2/3
white	1	2/4 = 1/2
	2	2/4 = 1/2

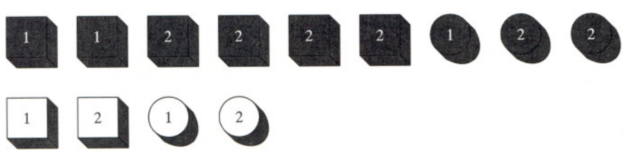


Figure 4: 13 objects with different shape, color, and label [Nea03, p. 8].

Chain rule

Lemma 2 (Chain rule). *Let* $X_1, X_2, \ldots, X_n$ *be variables. Then*

$$p(X_1, X_2, \ldots, X_n) = p(X_n|X_1, \ldots, X_{n-1}) \cdots p(X_2|X_1) \cdot p(X_1)$$

Potentials

Let

$$\mathcal{X} = \{X_1, \ldots, X_n\}$$

be a set of sets. We call any map

$$q : X_1 \times \cdots \times X_n \to \mathbb{R}_0^+$$

a **potential** on $\mathcal{X}$ and $\text{dom}(q) := \mathcal{X}$ its **set of domains.**

A potential q can be described as n-dimensional tensor indexed by the elements of the sets X_i.

Example 11. Let p be a joint probability distribution of a set $\mathcal{X}$ of random variables, i.e.,

$$p : X_1 \times \cdots \times X_n \to [0, 1]$$

Then p is a potential with domain $\mathcal{X}$.

$X_1 = P =$ Pain	Y				N			
$X_2 = W =$ Weightloss	Y		N		Y		N	
$X_3 = V =$ Vomiting	Y	N	Y	N	Y	N	Y	N
$X_4 = A =$ Adeno Y	0.220	0.220	0.025	0.025	0.095	0.095	0.010	0.010
N	0.004	0.009	0.005	0.012	0.031	0.076	0.050	0.113

Figure 5: Joint probability distribution of four random variables X_1 (pain), X_2 (weightloss), X_3 (vomiting) and X_4 (adeno) as potential.

Multiplication of potentials

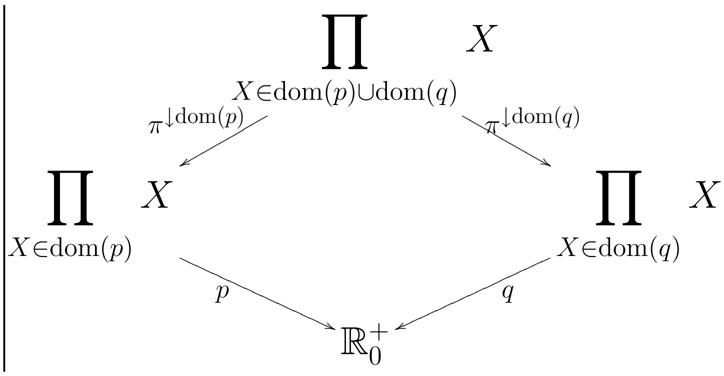
Let p, q be two potentials. We define

$$\begin{aligned}(p \cdot q) : \prod_{X \in \text{dom}(p) \cup \text{dom}(q)} X &\to \mathbb{R}_0^+ \\ x &\mapsto p(\pi^{\downarrow \text{dom}(p)}(x)) \cdot q(\pi^{\downarrow \text{dom}(q)}(x))\end{aligned}$$

as the **(outer) product of** p **and** q, where

$$\pi^{\downarrow \text{dom}(p)} : \prod_{X \in \text{dom}(p) \cup \text{dom}(q)} X \to \prod_{X \in \text{dom}(p)} X$$

is the canonical projection.



Multiplication of potentials / examples (1/2)

Example 12. Let

$$p := \begin{pmatrix} 0.2 \\ 0.3 \\ 0.5 \end{pmatrix}, \quad q := \begin{pmatrix} 0.1 \\ 0.2 \\ 0.4 \\ 0.3 \end{pmatrix}$$

be two vectors ("one-dimensional potentials"). Then

$$p \cdot q := \begin{pmatrix} 0.2 \cdot 0.1 & 0.2 \cdot 0.2 & 0.2 \cdot 0.4 & 0.2 \cdot 0.3 \\ 0.3 \cdot 0.1 & 0.3 \cdot 0.2 & 0.3 \cdot 0.4 & 0.3 \cdot 0.3 \\ 0.5 \cdot 0.1 & 0.5 \cdot 0.2 & 0.5 \cdot 0.4 & 0.5 \cdot 0.3 \end{pmatrix}$$

is their usual outer product.

Let

$$r := \begin{pmatrix} 0.1 \\ 0.2 \\ 0.4 \end{pmatrix}$$

be a third vector over the same domain as p, then

$$p \cdot r := \begin{pmatrix} 0.2 \cdot 0.1 \\ 0.3 \cdot 0.2 \\ 0.5 \cdot 0.4 \end{pmatrix}$$

is their element-wise product.

Multiplication of potentials / examples (1/2)

Example 13. Let

$$p := \begin{pmatrix} 0.2 & 0.1 \\ 0.3 & 0.2 \\ 0.5 & 0.4 \end{pmatrix}, \quad q := \begin{pmatrix} 0.1 & 0.6 \\ 0.2 & 0.1 \\ 0.4 & 0.3 \end{pmatrix}$$

be two matrices ("two-dimensional potentials") over the same domains. Then

$$p \cdot q := \begin{pmatrix} 0.2 \cdot 0.1 & 0.1 \cdot 0.6 \\ 0.3 \cdot 0.2 & 0.2 \cdot 0.1 \\ 0.5 \cdot 0.4 & 0.4 \cdot 0.3 \end{pmatrix}$$

is their element-wise product.

Let

$$r := \begin{pmatrix} 0.1 & 0.2 & 0.4 & 0.3 \\ 0.6 & 0.1 & 0.1 & 0.2 \end{pmatrix}$$

be a third matrix that has only one domain in common with p. Then $p \cdot r$ is a three-dimensional potential, e.g.,

$$(p \cdot r)_{3,2,4} = p_{3,2} \cdot r_{2,4} = 0.4 \cdot 0.2$$

Marginalization of potentials

Definition 10. Let p be a potential and $\mathcal{Y} \subseteq \mathrm{dom}(p)$ a subset of its domain. We define

$$p^{\downarrow \mathcal{Y}} : \prod_{X \in \mathcal{Y}} X \rightarrow \mathbb{R}_0^+$$
$$x \mapsto \sum_{x' \in \prod_{X \in \mathrm{dom}(p) \setminus \mathcal{Y}} X} p(\iota(x, x'))$$

as the **projection of p down to $\mathcal{Y}$** (or as **marginalization p out of $\mathrm{dom}(p) \setminus \mathcal{Y}$**) where

$$\iota : (\prod_{X \in \mathcal{Y}} X) \times (\prod_{X \in \mathrm{dom}(p) \setminus \mathcal{Y}} X) \rightarrow \prod_{X \in \mathrm{dom}(p)} X$$

is the canonical bijection.

Example 14. Assume the joint probability distribution of four random variables P (pain), W (weightloss), V (vomiting) and A (adeno) given in fig. 5 as potential p.

If we project p down to V and A, we get the potential $p^{\downarrow V,A}$:

Vomiting	Y	N
Adeno Y	0.350	0.350
N	0.090	0.210

$X_1 = P =$ Pain	Y				N			
$X_2 = W =$ Weightloss	Y		N		Y		N	
$X_3 = V =$ Vomiting	Y	N	Y	N	Y	N	Y	N
$X_4 = A =$ Adeno Y	0.220	0.220	0.025	0.025	0.095	0.095	0.010	0.010
N	0.004	0.009	0.005	0.012	0.031	0.076	0.050	0.113

Conditioning of potentials

Definition 11. By $p > 0$ we mean

$$p(x) > 0, \quad \text{for all } x \in \prod \mathrm{dom}(p)$$

Then p is called **non-extreme**.

For two potentials p, q with $q > 0$, by p/q we mean $p \cdot q^{-1}$ where

$$q^{-1}(y) := \frac{1}{q(y)}, \quad \text{for all } y \in \prod \mathrm{dom}(q)$$

For a potential p and a subset $\mathcal{Y} \subseteq \mathrm{dom}(p)$ of its domains with $p^{\downarrow \mathcal{Y}} > 0$ we define

$$p^{|\mathcal{Y}} := \frac{p}{p^{\downarrow \mathcal{Y}}}$$

as **conditioning of p at $\mathcal{Y}$**.

A potential conditioned at $\mathcal{Y}$ sums to 1 for all fixed values of $\mathcal{Y}$, i.e.,

$$(p^{|\mathcal{Y}})^{\downarrow \mathcal{Y}} \equiv 1$$

Example 15. Let p be the potential

$$p := \begin{pmatrix} 0.4 & 0.1 \\ 0.2 & 0.3 \end{pmatrix}$$

on two variables R (rows) and C (columns) with the domains $\mathrm{dom}(R) = \mathrm{dom}(C) = \{1, 2\}$.

If we conditioning on C we get

$$p := \begin{pmatrix} 2/3 & 1/4 \\ 1/3 & 3/4 \end{pmatrix}$$

i.e., if p is a joint probability distribution, we get the conditional probability distribution $p(R|C)$.

Conditioning of potentials / example

Example 16. If q is another potential

$$p := \begin{pmatrix} 80 & 20 \\ 40 & 60 \end{pmatrix}$$

that is not a joint probability distribution, we can **normalize** q by conditioning on $\emptyset$. Here

$$q^{|\emptyset} = p = \begin{pmatrix} 0.4 & 0.1 \\ 0.2 & 0.3 \end{pmatrix}$$

Chain rule revisited

Lemma 3 (chain rule)**.** *Let p be a potential and $\mathcal{Y} \subseteq \mathrm{dom}(p)$ a subset of its domains with $p^{\downarrow \mathcal{Y}} > 0$. Then*

$$p = p^{|\mathcal{Y}} \cdot p^{\downarrow \mathcal{Y}}$$

Let

$$\mathcal{Y}_1 \subset \mathcal{Y}_2 \subset \cdots \subset \mathcal{Y}_{m-1} \subset \mathcal{Y}_m = \mathrm{dom}(p)$$

be a sequence of subsets of $\mathrm{dom}(p)$ with $p^{\downarrow \mathcal{Y}_i} > 0$ for all i. Then

$$p = p^{\downarrow \mathcal{Y}_1} \prod_{i=1}^{m-1} p^{\downarrow \mathcal{Y}_{i+1} | \mathcal{Y}_i}$$
$$= p^{|\mathcal{Y}_{m-1}} \cdot p^{\downarrow \mathcal{Y}_{m-1} | \mathcal{Y}_{m-2}} \cdots p^{\downarrow \mathcal{Y}_2 | \mathcal{Y}_1} \cdot p^{\downarrow \mathcal{Y}_1}$$

Example 17. If p is a probability distribution over the variables $\mathrm{dom}(x) = \{X_1, \ldots, X_n\}$,

$$\mathcal{Y}_i := \{X_1, \ldots, X_i\}$$

and all marginals $p^{\downarrow X_1, \ldots, X_i} > 0$ (e.g., $p > 0$).
We write

$$p(X_i | X_1, \ldots, X_{i-1}) := p^{\downarrow X_1, \ldots, X_i | X_1, \ldots X_{i-1}}$$
$$= p^{\downarrow \mathcal{Y}_i | \mathcal{Y}_{i-1}}$$

Then the chain rule can be written as

$$p(X_1, X_2, \ldots, X_n) = p(X_n | X_1, \ldots, X_{n-1}) \cdots p(X_2 | X_1) \cdot p(X_1)$$

Variable independence revisited

Definition 12. Let p be a potential and $\mathcal{X}, \mathcal{Y} \subseteq \mathrm{dom}(p)$ be two subsets of its domains. We call $\mathcal{X}$ and $\mathcal{Y}$ **independent**, if

$$p^{\downarrow \mathcal{X} \cup \mathcal{Y}} = p^{\downarrow \mathcal{X}} \cdot p^{\downarrow \mathcal{Y}}$$

Let $\mathcal{Z} \subseteq \mathrm{dom}(p)$ a third subset of its domains. Then $\mathcal{X}$ and $\mathcal{Y}$ are called **conditionally independent given** $\mathcal{Z}$, if

$$p^{\downarrow \mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z}} \cdot p^{\downarrow \mathcal{Z}} = p^{\downarrow \mathcal{X} \cup \mathcal{Z}} \cdot p^{\downarrow \mathcal{Y} \cup \mathcal{Z}}$$

or equivalently

$$p^{\downarrow \mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z} | \mathcal{Y} \cup \mathcal{Z}} = p^{\downarrow \mathcal{X} \cup \mathcal{Z} | \mathcal{Z}} \cdot 1_{\mathcal{Y}}$$

(for all $x \in \prod \mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z}$ with $p^{\downarrow \mathcal{Y} \cup \mathcal{Z}}(\pi^{\downarrow \mathcal{Y} \cup \mathcal{Z}}(x)) > 0$).

1. Basic Probability Calculus

2. Tensor calculus for conditional probabilities

3. Separation in undirected graphs

Graphs

Definition 13. Let V be any set and

$$E \subseteq \mathcal{P}^2(V) := \{\{x, y\} \mid x, y \in V\}$$

be a subset of sets of unordered pairs of V. Then $G := (V, E)$ is called an **undirected graph**. The elements of V are called **vertices** or **nodes**, the elements of E **edges**.

Let $e = \{x, y\} \in E$ be an edge, then we call the vertices x, y **incident** to the edge e. We call two vertices $x, y \in V$ **adjacent**, if there is an edge $\{x, y\} \in E$.

The set of all vertices adjacent with a given vertex $x \in V$ is called its **fan**:

$$\text{fan}(x) := \{y \in V \mid \{x, y\} \in E\}$$

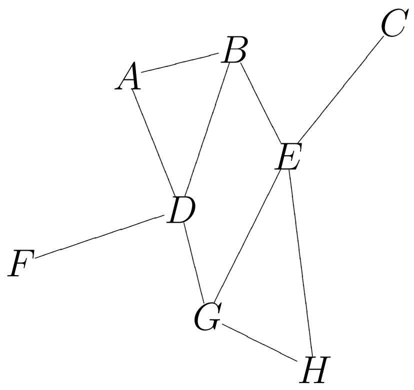


Figure 7: Example graph.

Paths on graphs

Definition 14. Let V be a set. We call $V^* := \bigcup_{i \in \mathbb{N}} V^i$ the **set of finite sequences in** V. The length of a sequence $s \in V^*$ is denoted by $|s|$.

Let $G = (V, E)$ be a graph. We call

$$G^* := V^*_{|G} := \{p \in V^* \mid \{p_i, p_{i+1}\} \in E, \\ i = 1, \ldots, |p| - 1\}$$

the **set of paths on** G.

Any contiguous subsequence of a path $p \in G^*$ is called a **subpath of** p, i.e. any path $(p_i, p_{i+1}, \ldots, p_j)$ with $1 \leq i \leq j \leq n$. The subpath $(p_2, p_3, \ldots, p_{n-1})$ is called the **interior of** p. A path of length $|p| \geq 2$ is called **proper**.

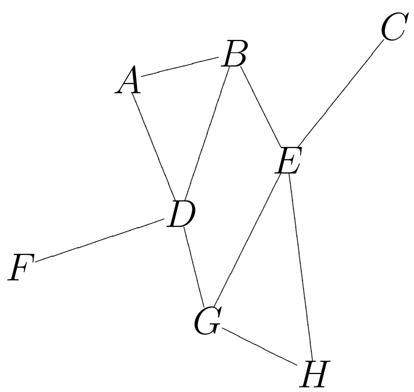


Figure 8: Example graph.

The sequences

$$(A, D, G, H)$$
$$(C, E, B, D)$$
$$(F)$$

are paths on G, but the sequences

$$(A, D, E, C)$$
$$(A, H, C, F)$$

are not.

Separation in graphs (u-separation)

Definition 15. Let $G := (V, E)$ be a graph. Let $Z \subseteq V$ be a subset of vertices. We say, two vertices $x, y \in V$ are **separated by** Z **in** G, if every path from x to y contains some vertex of Z ($\forall p \in G^* : p_1 = x, p_{|p|} = y \Rightarrow \exists i \in \{1, \ldots, n\} : p_i \in Z$).

Let $X, Y, Z \subseteq V$ be three disjoint subsets of vertices. We say, the vertices X and Y are **separated by** Z **in** G, if every path from any vertex from X to any vertex from Y is separated by Z, i.e., contains some vertex of Z.

We write $I_G(X, Y|Z)$ for the statement, that X and Y are separated by Z in G. I_G is an example for a ternary relation on $\mathcal{P}(V)$. We call I_G the **u-separation relation in** G.

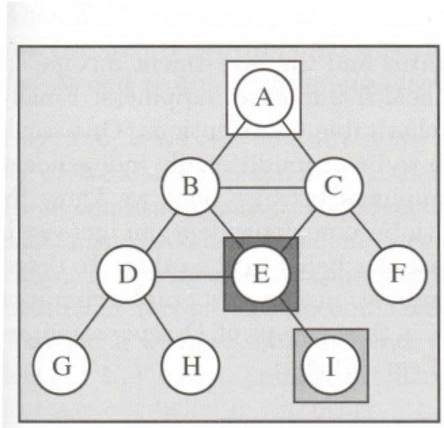


Figure 9: Example for u-separation [CGH97, p. 179].

Separation in graphs (u-separation)

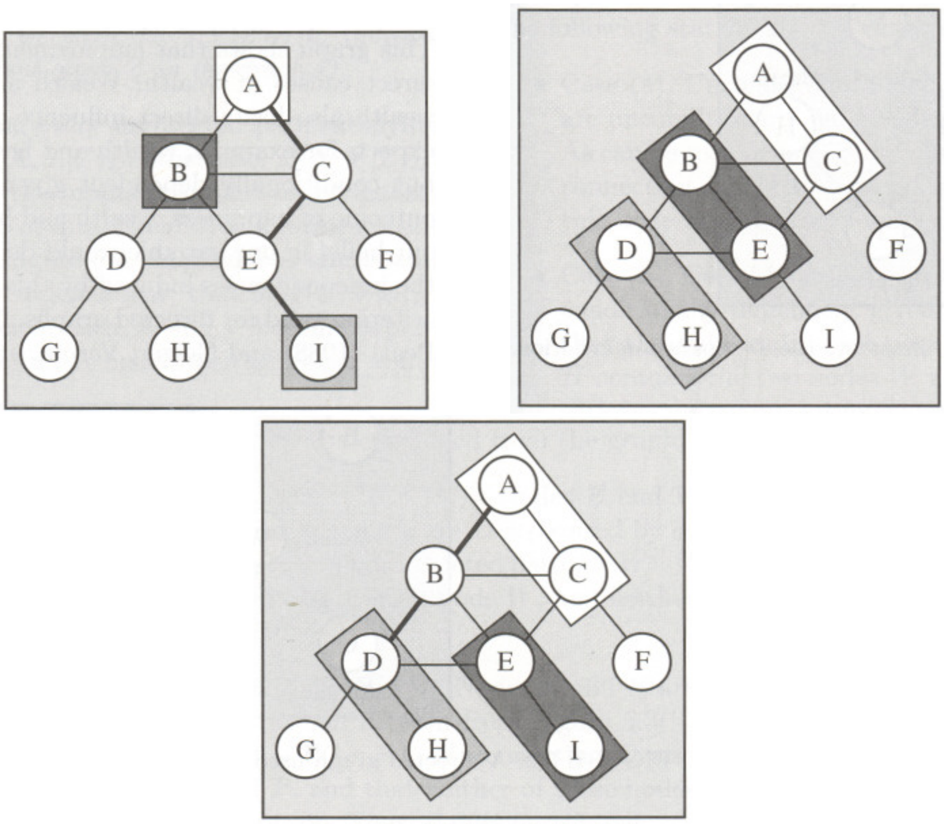


Figure 10: More examples for u-separation [CGH97, p. 179].

Properties of ternary relations

Definition 16. Let V be any set and I a ternary relation on $\mathcal{P}(V)$, i.e., $I \subseteq (\mathcal{P}(V))^3$.

I is called **symmetric**, if

$$I(X, Y|Z) \Rightarrow I(Y, X|Z)$$

I is called **(right-)decomposable**, if

$$I(X, Y|Z) \Rightarrow I(X, Y'|Z) \quad \text{for any } Y' \subseteq Y$$

I is called **(right-)composable**, if

$$I(X, Y|Z) \text{ and } I(X, Y'|Z) \Rightarrow I(X, Y \cup Y'|Z)$$

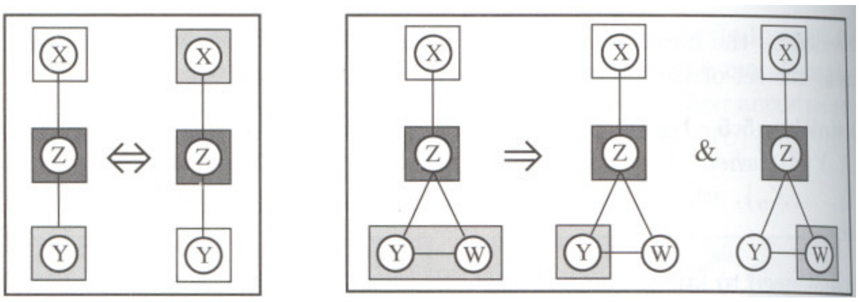


Figure 11: Examples for a) symmetry and b) decomposition [CGH97, p. 186].

Properties of ternary relations

Definition 17. I is called **strongly unionable**, if

$$I(X, Y|Z) \Rightarrow I(X, Y|Z \cup Z') \quad \text{for all } Z' \text{ disjunct with } X, Y$$

I is called **(right-)weakly unionable**, if

$$I(X, Y|Z) \Rightarrow I(X, Y'|(Y \setminus Y') \cup Z) \quad \text{for any } Y' \subseteq Y$$

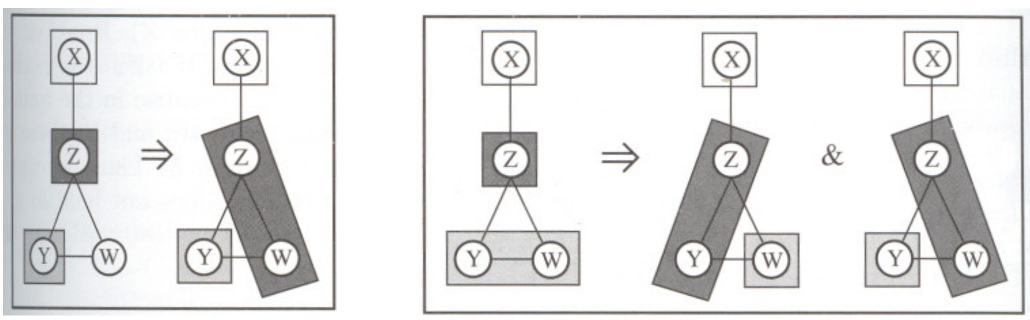


Figure 12: Examples for a) strong union and b) weak union [CGH97, p. 186,189].

Properties of ternary relations

Definition 18. I is called **(right-)contractable**, if

$$I(X,Y|Z) \text{ and } I(X,Y'|Y \cup Z) \Rightarrow I(X,Y \cup Y'|Z)$$

I is called **(right-)intersectable**, if

$$I(X,Y|Y' \cup Z) \text{ and } I(X,Y'|Y \cup Z) \Rightarrow I(X,Y \cup Y'|Z)$$

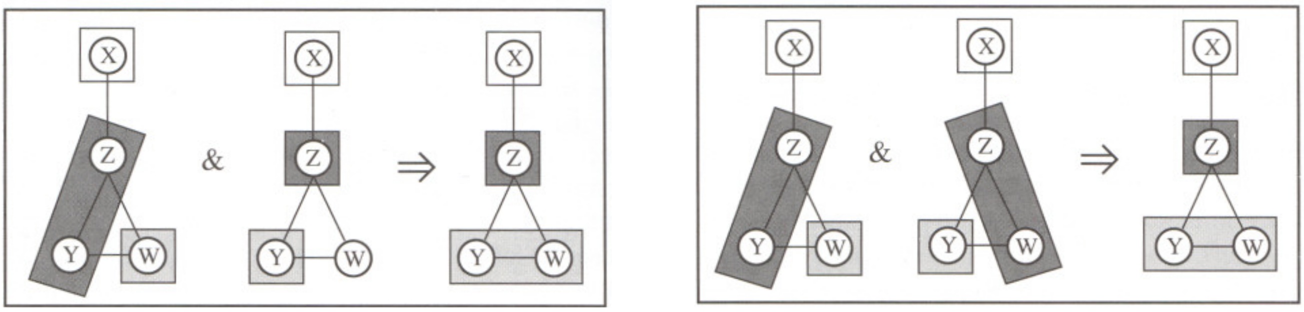


Figure 13: Examples for a) contraction and b) intersection [CGH97, p. 186].

Properties of ternary relations

Definition 19. I is called **strongly transitive**, if

$$I(X,Y|Z) \Rightarrow I(X,\{v\}|Z) \text{ or } I(\{v\},Y|Z) \quad \forall v \in V \setminus Z$$

I is called **weakly transitive**, if

$$I(X,Y|Z) \text{ and } I(X,Y|Z \cup \{v\}) \Rightarrow I(X,\{v\}|Z) \text{ or } I(\{v\},Y|Z) \quad \forall v \in V \setminus Z$$

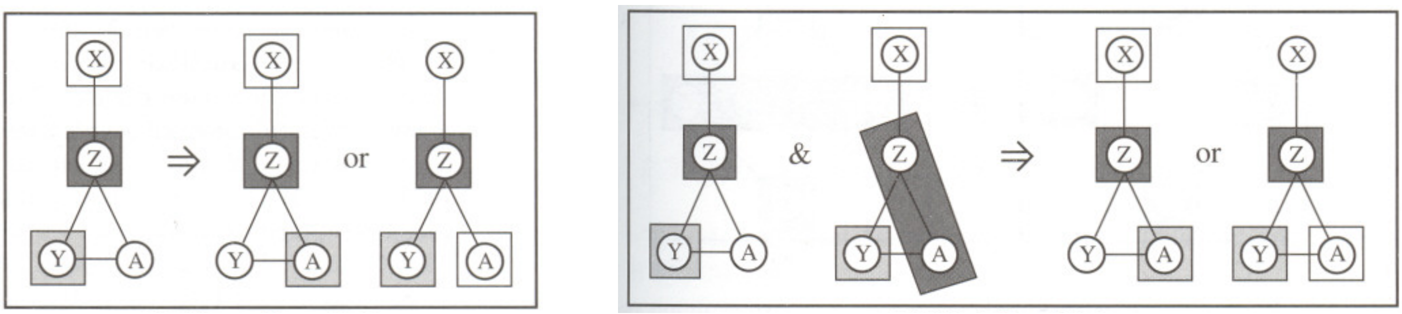


Figure 14: Examples for a) strong transitivity and b) weak transitivity. [CGH97, p. 189]

Properties of ternary relations

Definition 20. I is called **chordal**, if

$$I(\{a\},\{c\}|\{b,d\}) \text{ and } I(\{b\},\{d\}|\{a,c\}) \Rightarrow I(\{a\},\{c\}|\{b\}) \text{ or } I(\{a\},\{c\}|\{d\})$$

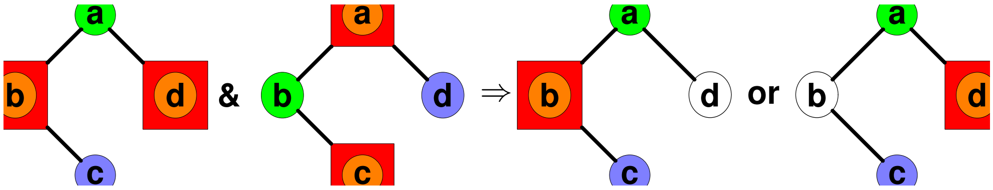


Figure 15: Example for chordality.

Properties of u-separation / no chardality

For u-separation the chordality property does not hold (in general).

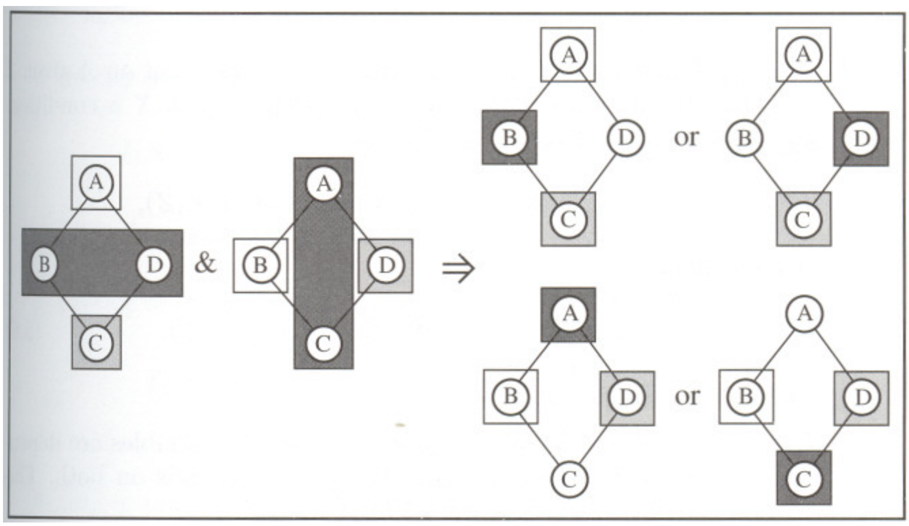


Figure 16: Counterexample for chordality in undirected graphs (u-separation) [CGH97, p. 189].

Properties of u-separation

relation	symmetry	decomposition	composition	strong union	weak union	contraction	intersection	strong transitivity	weak transitivity	chordality
u-separation	+	+	+	+	+	+	+	+	+	−

Checking u-separation

To test, if for a given graph $G = (V, E)$ two given sets $X, Y \subseteq V$ of vertices are u-separated by a third given set $Z \subseteq V$ of vertices, we may use standard breadth-first search to compute all vertices that can be reached from X (see, e.g., [OW02], [CLR90]).

```
breadth-first search(G, X) :
border := X
reached := ∅
while border ≠ ∅ do
        reached := reached ∪ border
        border := fan_G(border) \ reached
od
return reached
```

Figure 17: Breadth-first search algorithm for enumerating all vertices reachable from X.

For checking u-separation we have to tweak the algorithm

1. not to add vertices from Z to the border and
2. to stop if a vertex of Y has been reached.

```
check-u-separation(G, X, Y, Z) :
border := X
reached := ∅
while border ≠ ∅ do
        reached := reached ∪ border
        border := fan_G(border) \ reached \ Z
        if border ∩ Y ≠ ∅
            return false
        fi
od
return true
```

Figure 18: Breadth-first search algorithm for checking u-separation of X and Y by Z.

References

[BK02] Christian Borgelt and Rudolf Kruse. *Graphical Models*. Wiley, New York, 2002.

[CGH97] Enrique Castillo, José Manuel Gutiérrez, and Ali S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, New York, 1997.

[CLR90] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms*. MIT Press, Cambridge, Massachusetts, 1990.

[Nea03] Richard E. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, 2003.

[OW02] Thomas Ottmann and Peter Widmayer. *Algorithmen und Datenstrukturen*. Spektrum Verlag, Heidelberg, 2002.