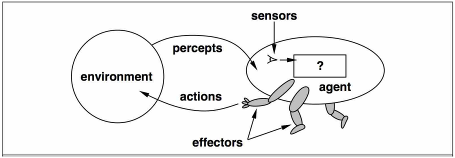
Knowledge Representation and Inference under Uncertainty: Bayesian Networks

Rational agents: acting under uncertainty



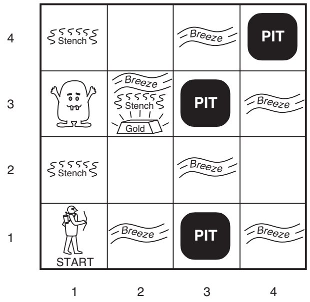
Examples

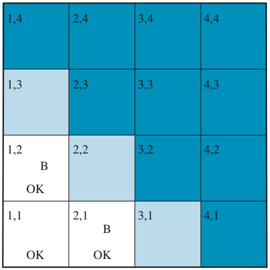
- medical diagnosis from patients' symptoms and medical tests' outcomes
- speech/image recognition from noisy audio/video signals
- self-driving vehicles: incomplete and noisy sensory data

Limitations of logical agents

Propositions: true, false, or unknown – no **degree of belief**

Example: in the wumpus world sensors report only **local** information





Knowledge about pits?

Limitations of logical agents

Propositions: true, false, or unknown – no **degree of belief**

Example: dental diagnosis

- toothache is caused by a cavity:

$$\forall p \; Symptom(p, Toothache) \Rightarrow Disease(p, Cavity)$$

- toothache is caused by a cavity, or an abscess, or...:

$$\forall p \; Symptom(p, Toothache) \Rightarrow$$
$$Disease(p, Cavity) \lor Disease(p, Abscess) \lor \ldots$$

- cavity causes toothache:

$$\forall p \; Disease(p, Cavity) \Rightarrow Symptom(p, Toothache)$$

Making rational decisions under uncertainty

Dealing with uncertainty: assigning a **degree of belief** to sentences of interest

Main tool: **probability theory** (degree of belief in $[0, 1]$)

Effective decision-making: **preferences** about the possible **outcomes** of actions – **utility theory**

Rational decision-making under uncertainty: **decision theory**

$$\textit{probability theory} + \textit{utility theory}$$

Probabilistic modelling and inference

Random variables

A description of the "state of the world" of interest to the agent

- notation: symbols with uppercase initial, e.g.:
  M, X, $Weather$, $Toothache$
- domain
  - Boolean, e.g.: $Toothache \in \{\mathrm{true}, \ \mathrm{false}\}$
  - discrete, e.g.: $Weather \in \{\mathrm{sunny}, \ \mathrm{rainy}, \ \mathrm{cloudy}, \ \mathrm{snow}\}$, $M \in \mathbb{N}$
  - continuous, e.g.: $X \in \mathbb{R}$

Probability distribution function (PDF)

Example: $\mathit{Weather} \in \langle \mathrm{sunny}, \mathrm{rain}, \mathrm{cloudy}, \mathrm{snow} \rangle$

$$
\begin{aligned}
P(\mathit{Weather} = \mathrm{sunny}) &= 0.7 \\
P(\mathit{Weather} = \mathrm{rain}) &= 0.2 \\
P(\mathit{Weather} = \mathrm{cloudy}) &= 0.08 \\
P(\mathit{Weather} = \mathrm{snow}) &= 0.02
\end{aligned}
$$

Vector notation: $\mathbf{P}(\mathit{Wheater}) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$

Axioms and theorems of probability theory:

- $0 \leq P(\cdot) \leq 1$ (axiom)
- $\sum_{x \in \mathcal{D}(X)} P(X = x) = 1$ (theorem)

(Full) joint PDF

Example: Boolean random variables describing a dentist's patient

- *Toothache* (the patient has a toothache)
- *Cavity* (the patient has a cavity)
- *Catch* (the dentist's steel probe catches in a tooth)

(Full) joint PDF $\mathbf{P}(Toothache, Cavity, Catch)$:

	$Toothache = t$		$Toothache = f$	
	$Catch = t$	$Catch = f$	$Catch = t$	$Catch = f$
$Cavity = t$	0.108	0.012	0.072	0.008
$Cavity = f$	0.016	0.064	0.144	0.576

Theorem:

$$\sum_{a,b,c \in \{\text{true},\text{false}\}} P(Toothache = a, Cavity = b, Catch = c) = 1$$

Events

Event: any combination of values of (a subset of) the random variables, e.g.:

$$\mathit{Cavity} = \mathit{true} \ \lor \ \mathit{Toothache} = \mathit{false}$$

Atomic event: any combination of values of **all** the random variables (a **complete** description of the "state of the world"), e.g.:

$$\mathit{Cavity} = \mathrm{true} \ \land \ \mathit{Toothache} = \mathrm{false} \ \land \mathit{Catch} = \mathrm{true}$$

- **mutually esclusive** and **exhaustive**
- any event is a **disjunction** of atomic events, e.g.:

$$\mathit{Cavity} = \mathrm{true} \ \lor \ \mathit{Toothache} = \mathrm{false} \equiv$$
$$(\mathit{Cavity} = \mathrm{true} \ \land \ \mathit{Toothache} = \mathrm{true} \ \land \mathit{Catch} = \mathrm{true}) \lor \ldots$$

Probabilistic inference

Computing the probability of an event of interest

A dentist may be interested in

- $P(\mathit{Cavity} = \text{true})$
- $P(\mathit{Cavity} = \text{true} \wedge \mathit{Toothache} = \text{true})$

Marginal probability: joint probability of any **subset** of random variables

Probabilistic inference: marginalisation (sum rule)

(Full) joint PDF $\mathbf{P}(\mathit{Toothache}, \mathit{Cavity}, \mathit{Catch})$:

	Toothache = *t*		*Toothache* = *f*	
	Catch = *t*	*Catch* = *f*	*Catch* = *t*	*Catch* = *f*
Cavity = *t*	0.108	0.012	0.072	0.008
Cavity = *f*	0.016	0.064	0.144	0.576

Marginal PDF $\mathbf{P}(\mathit{Cavity}, \mathit{Toothache})$:

		Tootache	
		true	false
Cavity	true	0.120	0.080
	false	0.080	0.720

Marginalisation, or sum rule:

$$\mathbf{P}(X_1, \ldots, X_p) = \sum_{x_{p+1} \in \mathcal{X}_{p+1}, \ldots, x_n \in \mathcal{X}_n} \mathbf{P}(X_1, \ldots, X_p, X_{p+1} = x_{p+1}, \ldots, X_n = x_n)$$

Prior and posterior (conditional) probability

A dentist may be also interested in posterior (conditional) probabilities

- $P(\mathit{Cavity} = \mathrm{true} | \mathit{Toothache} = \mathrm{true}) \triangleq \frac{P(\mathit{Cavity}=\mathrm{true}, \mathit{Toothache}=\mathrm{true})}{P(\mathit{Toothache}=\mathrm{true})}$
- $P(\mathit{Cavity} = \mathrm{true} | \mathit{Catch} = \mathrm{true}) \triangleq \frac{P(\mathit{Cavity}=\mathrm{true}, \mathit{Catch}=\mathrm{true})}{P(\mathit{Catch}=\mathrm{true})}$
- $P(\mathit{Cavity} = \mathrm{true} | \mathit{Toothache} = \mathrm{true}, \mathit{Catch} = \mathrm{false}) \triangleq$
  $\frac{P(\mathit{Cavity}=\mathrm{true}, \mathit{Toothache}=\mathrm{true}, \mathit{Catch}=\mathrm{false})}{P(\mathit{Toothache}=\mathrm{true}, \mathit{Catch}=\mathrm{false})}$
- $P(\mathit{Cavity} = \mathrm{true}, \mathit{Catch} = \mathrm{false} | \mathit{Toothache} = \mathrm{true}) \triangleq$
  $\frac{P(\mathit{Cavity}=\mathrm{true}, \mathit{Catch}=\mathrm{false}, \mathit{Toothache}=\mathrm{true})}{P(\mathit{Toothache}=\mathrm{true})}$

Conditional PDF

Conditional probability: $P(Cavity = \mathrm{true} | Toothache = \mathrm{true})$

Conditional PDF: $\mathbf{P}(Cavity | Toothache = \mathrm{true}) = \langle 0.8, 0.2 \rangle$

Also the values of a conditional PDF sum to 1

Every value of the conditioning event corresponds to a **distinct** conditional PDF, e.g.:

- $\mathbf{P}(Cavity | Toothache = \mathrm{true}) = \langle 0.8, 0.2 \rangle$
- $\mathbf{P}(Cavity | Toothache = \mathrm{false}) = \langle 0.05, 0.95 \rangle$

Probabilistic inference

Also **conditional** probabilities can be computed from the **full joint** PDF

$$
\begin{aligned}
&P(\mathit{Cavity} = \mathrm{true} | \mathit{Toothache} = \mathrm{true}) \\
&= \frac{P(\mathit{Cavity} = \mathrm{true}, \mathit{Toothache} = \mathrm{true})}{P(\mathit{Toothache} = \mathrm{true})} \text{ (by definition)}
\end{aligned}
$$

Full joint PDF:

	$\mathit{Toothache} = \mathrm{t}$		$\mathit{Toothache} = \mathrm{t}$	
	$\mathit{Catch} = \mathrm{t}$	$\mathit{Catch} = \mathrm{f}$	$\mathit{Catch} = \mathrm{t}$	$\mathit{Catch} = \mathrm{f}$
$\mathit{Cavity} = \mathrm{t}$	0.108	0.012	0.072	0.008
$\mathit{Cavity} = \mathrm{f}$	0.016	0.064	0.144	0.576

Sum rule:

$$
\begin{aligned}
&P(\mathit{Cavity} = \mathrm{t}, \mathit{Toothache} = \mathrm{t}) = 0.108 + 0.012 = 0.120 \\
&P(\mathit{Toothache} = \mathrm{t}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.200
\end{aligned}
$$

Probabilistic inference: summary

If the full joint PDF of random variables $X_1, \dots, X_n$ is **known**

- marginal (prior) PDF: sum rule

$$\mathbf{P}(X_1, \dots, X_p) = \sum_{x_{p+1} \in \mathcal{X}_{p+1}, \dots, x_n \in \mathcal{X}_n} \mathbf{P}(X_1, \dots, X_p, X_{p+1} = x_{p+1}, \dots, X_n = x_n)$$

- conditional (posterior) PDF: definition + sum rule

$$\mathbf{P}(X_1, \dots, X_p | X_{p+1}, \dots, X_q) = \frac{\mathbf{P}(X_1, \dots, X_p, X_{p+1}, \dots, X_q)}{\mathbf{P}(X_{p+1}, \dots, X_q)} = \dots$$

Probabilistic inference using the full joint PDF: issues

How to compute or estimate the full joint PDF?

- $Coin \in \langle \text{heads}, \text{tails} \rangle : \quad \mathbf{P}(Coin) = \langle ?, ? \rangle$
- $Dice \in \langle ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ \rangle : \quad \mathbf{P}(Dice) = \langle ?, ?, ?, ?, ?, ? \rangle$

Classical (a priori) probability: mutually exclusive, equally likely, random atomic events

- $Coin \in \langle \text{heads}, \text{tails} \rangle : \quad \mathbf{P}(Coin) = \langle \frac{1}{2}, \frac{1}{2} \rangle$
- $Dice \in \langle ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ \rangle : \quad \mathbf{P}(Dice) = \langle \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \rangle$

Same approach for the probability of **any** related event, e.g.:

- getting an even face up (⚁ or ⚃ or ⚅)
- getting a sum of the faces up equal to 6 after throwing two dice (⚀, ⚄ or ⚁, ⚃, . . .)

Probabilistic inference using the full joint PDF: issues

How to compute or estimate the full joint PDF?

	Toothache = t		*Toothache* = f	
	Catch = t	*Catch* = f	*Catch* = t	*Catch* = f
Cavity = t	?	?	?	?
Cavity = f	?	?	?	?

Frequentist (a posteriori) probability: when events of interest can be observed under **similar** and **uniform** conditions

Examples

- getting ⚅ after throwing a **loaded** dice
- a chip manufactured by company *XYZ* is faulty
- a student in CECAI graduates within 2 years

Probabilistic inference using the full joint PDF: issues

How to compute or estimate the full joint PDF?

- the piano player John Smith will break one or both of his hands within the next 10 years
- the third World War will start within 2026

Subjective probability (e.g., domain experts' judgement)

Probabilistic inference using the full joint PDF: issues

Effort required to estimate the full joint PDF

Example: full joint PDF of n Boolean random variables

	$Toothache = \mathrm{t}$		$Toothache = \mathrm{f}$	
	$Catch = \mathrm{t}$	$Catch = \mathrm{f}$	$Catch = \mathrm{t}$	$Catch = \mathrm{f}$
$Cavity = \mathrm{t}$	?	?	?	?
$Cavity = \mathrm{f}$	?	?	?	?

$2^n - 1$ probability values must be specified; many of them may be not easy to estimate, too

Probabilistic inference using the full joint PDF: issues

Computational complexity of inference

Example: inference over n Boolean random variables:

$$\mathbf{P}(X_1, \ldots, X_p) = \text{(sum rule)}$$
$$\sum_{x_{p+1} \in \mathcal{X}_{p+1}, \ldots, x_n \in \mathcal{X}_n} \mathbf{P}(X_1, \ldots, X_p, X_{p+1} = x_{p+1}, \ldots, X_n = x_n)$$

sum of 2^n probability values (2^{n-p} values for **each** of the 2^p values of $X_1, \ldots, X_p$)

Simplifying probabilistic modelling and inference

A useful tool: the product rule

$$\underbrace{\mathbf{P}(X|Y) \triangleq \frac{\mathbf{P}(X,Y)}{\mathbf{P}(Y)}}_{\text{definition of conditional probability}} \quad \Rightarrow \quad \underbrace{\mathbf{P}(X,Y) = \mathbf{P}(X|Y)\mathbf{P}(Y)}_{\text{product rule}}$$

Extension to **groups** of random variables, e.g.:

$$\mathbf{P}(X_1,\ldots,X_q) = \mathbf{P}(X_1,\ldots,X_p|X_{p+1},\ldots,X_q)\mathbf{P}(X_{p+1},\ldots,X_q)$$

Extension to **conditional** probabilities, e.g.:

$$\mathbf{P}(X_1,X_2|X_3) = \mathbf{P}(X_1|X_2,X_3)\mathbf{P}(X_2|X_3)$$

Simplifying probabilistic modelling and inference

Besides the Boolean variables *Toothache*, *Catch* and *Cavity*, our dentist would like to use *Weather* $\in \langle \mathrm{sunny}, \mathrm{rain}, \mathrm{cloudy}, \mathrm{snow} \rangle$

How many probability values must be estimated to specify the full joint PDF $\mathbf{P}(\mathit{Toothache}, \mathit{Catch}, \mathit{Cavity}, \mathit{Weather})$?
$(2 \times 2 \times 2 \times 4) - 1 = 31 \ldots$

Is there any cause–effect relation between weather and dental problems?

Simplifying probabilistic modelling and inference

Using the product rule:

$$\mathbf{P}(Toothache, Catch, Cavity, Weather) =$$
$$\mathbf{P}(Toothache, Catch, Cavity | Weather)\mathbf{P}(Weather)$$

Weather can be **assumed** to be (statistically) independent on *Toothache*, *Catch* and *Cavity*:

$$\mathbf{P}(Toothache, Catch, Cavity | Weather) = \mathbf{P}(Toothache, Catch, Cavity)$$

It follows that:

$$\mathbf{P}(Toothache, Catch, Cavity, Weather) =$$
$$\mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

How many probability values must be estimated to specify $\mathbf{P}(Toothache, Catch, Cavity)$ and $\mathbf{P}(Weather)$?
$(2 \times 2 \times 2 - 1) + (4 - 1) = 10$!

Independence: formal definition

Two variables X and Y are independent, if and only if

$$\mathbf{P}(X|Y) = \mathbf{P}(X)$$

Equivalent conditions (notice the symmetry):

$$\begin{aligned} \mathbf{P}(Y|X) &= \mathbf{P}(Y) \\ \mathbf{P}(X,Y) &= \mathbf{P}(X)\mathbf{P}(Y) \end{aligned}$$

Extension to **groups** of random variables, e.g. (**equivalent** conditions):

$$\begin{aligned} \mathbf{P}(X_1,\ldots,X_p|X_{p+1},\ldots,X_q) &= \mathbf{P}(X_1,\ldots,X_p) \\ \mathbf{P}(X_{p+1},\ldots,X_q|X_1,\ldots,X_p) &= \mathbf{P}(X_{p+1},\ldots,X_q) \\ \mathbf{P}(X_1,\ldots,X_p,X_{p+1},\ldots,X_q) &= \mathbf{P}(X_1,\ldots,X_p)\mathbf{P}(X_1,\ldots,X_p) \end{aligned}$$

Why is independence useful?

Example: n Boolean variables

Full joint PDF $\mathbf{P}(X_1, \ldots, X_n)$: specified by $2^n - 1$ probability values...

...if the variables were **all** independent:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i)$$

only $n \times (2^1 - 1) = n$ probability values are required!

Unfortunately, absolute independence is rare in practice...

Simplifying probabilistic modelling and inference

A weaker form of independence can be exploited, considering the cause–effect relation between variables

Toothache and *Catch* are possible **effects** of *Cavity*: using the product rule, the joint PDF can be rewritten into a causal form $P(effects|cause)$:

$$\mathbf{P}(Toothache, Catch, Cavity) = \underbrace{\mathbf{P}(Toothache, Catch|Cavity)}_{P(effects|cause)}\mathbf{P}(Cavity)$$

What can be said about $\mathbf{P}(Toothache, Catch|Cavity)$?

Simplifying probabilistic modelling and inference

Note first that *Toothache* and *Catch* cannot be considered absolutely independent, i.e.:

$$\mathbf{P}(Toothache|Catch) \neq \mathbf{P}(Toothache)$$

Example:

$$\mathbf{P}(Toothache|Catch = \text{true}) \neq \mathbf{P}(Toothache|Catch = \text{false})$$

However, they can be **assumed** to be conditionally independent **given** *Cavity*:

$$\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$$

Simplifying probabilistic modelling and inference

Note that the expression:

$$\mathbf{P}(\mathit{Toothache}|\mathit{Catch}, \mathit{Cavity}) = \mathbf{P}(\mathit{Toothache}|\mathit{Cavity})$$

is **equivalent** to:

$$\begin{aligned}
\mathbf{P}(\mathit{Catch}|\mathit{Toothache}, \mathit{Cavity}) &= \mathbf{P}(\mathit{Catch}|\mathit{Cavity}) \\
\mathbf{P}(\mathit{Toothache}, \mathit{Catch}|\mathit{Cavity}) &= \mathbf{P}(\mathit{Toothache}|\mathit{Cavity})\mathbf{P}(\mathit{Catch}|\mathit{Cavity})
\end{aligned}$$

Conditional independence: formal definition

Two variables X and Y are conditionally independent **given** another variable Z, if and only if

$$\mathbf{P}(X|Y,Z) = \mathbf{P}(X|Z)$$

Equivalent conditions:

$$\begin{aligned}\mathbf{P}(Y|X,Z) &= \mathbf{P}(Y|Z)\\ \mathbf{P}(X,Y|Z) &= \mathbf{P}(X|Z)\mathbf{P}(Y|Z)\end{aligned}$$

Extension to **groups** of random variables (**equivalent** conditions):

$$\begin{aligned}\mathbf{P}(\mathbf{X}|\mathbf{Y},\mathbf{Z}) &= \mathbf{P}(\mathbf{X}|\mathbf{Z})\\ \mathbf{P}(\mathbf{Y}|\mathbf{X},\mathbf{Z}) &= \mathbf{P}(\mathbf{Y}|\mathbf{Z})\\ \mathbf{P}(\mathbf{X},\mathbf{Y}|\mathbf{Z}) &= \mathbf{P}(\mathbf{X}|\mathbf{Z})\mathbf{P}(\mathbf{Y}|\mathbf{Z})\end{aligned}$$

Usefulness of conditional independence

From the above conditional independence assumption it follows that:

$$\mathbf{P}(Toothache, Catch, Cavity) = \text{(product rule)}$$
$$\mathbf{P}(Toothache, Catch|Cavity)\mathbf{P}(Cavity) = \text{(conditional independence)}$$
$$\mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)$$

To specify the full joint distribution, $2^3 - 1 = 7$ probability values must be estimated...

To specify the other distributions one needs to estimate

- $\mathbf{P}(Toothache|Cavity)$: $2 \times (2^1 - 1) = 2$ values
- $\mathbf{P}(Catch|Cavity)$: $2 \times (2^1 - 1) = 2$ values
- $\mathbf{P}(Cavity)$: $2^1 - 1 = 1$ value

for a total of 5 values – a **negligible** gain?

Usefulness of conditional independence

Example: n Boolean variables

The conditional PDF

$$\mathbf{P}(X_1, \ldots, X_{n-1} | X_n)$$

is specified by $2 \times (2^{n-1} - 1) = 2^n - 2$ probability values...

...if $X_1, \ldots, X_{n-1}$ were conditionally independent given X_n:

$$\mathbf{P}(X_1, \ldots, X_n | Y) = \prod_{i=1}^{n-1} \mathbf{P}(X_i | X_n) \ ,$$

only $(n-1) \times 2 \times (2^1 - 1) = 2(n-1)$ probability values are required!

Simplifying probabilistic modelling and inference

Potential advantage of causal inference in the form $P(\mathit{effects}|\mathit{cause})$: distinct effects of a common cause may be conditionally independent, **given** the cause, e.g.:

$$\mathbf{P}(\underbrace{Toothache, Catch}_{\text{effects}} | \underbrace{Cavity}_{\text{cause}}) = \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$$

In practice, diagnostic inference $P(\mathit{cause}|\mathit{effects})$ is often required, e.g.:

$$P(Cavity|Toothache, Catch)$$

Unfortunately, causal knowledge is usually easier to obtain than diagnostic. . .

Diagnostic inference by means of causal inference

A useful tool: Bayes' rule

Two **equivalent** expressions of the product rule:

$$
\begin{aligned}
\mathbf{P}(X, Y) &= \mathbf{P}(X|Y)\mathbf{P}(Y) \\
\mathbf{P}(Y, X) &= \mathbf{P}(Y|X)\mathbf{P}(X)
\end{aligned}
$$

it follows that:

$$\underbrace{\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)}}_{\text{Bayes' rule}}$$

Also the denominator $\mathbf{P}(X)$ can be rewritten as a function of $\mathbf{P}(Y|X)$, through the sum and product rules:

$$\mathbf{P}(X) = \sum_{y \in \mathcal{Y}} \mathbf{P}(X, Y = y) = \sum_{y \in \mathcal{Y}} \mathbf{P}(X|Y = y)P(Y = y)$$

Diagnostic inference by means of causal inference

An example of the application of Bayes' rule:

$$\mathbf{P}(Cavity \mid Toothache, Catch) =$$

$$\frac{\mathbf{P}(Toothache, Catch \mid Cavity)\ \mathbf{P}(Cavity)}{\mathbf{P}(Toothache, Catch)} =$$

$$\frac{\mathbf{P}(Toothache, Catch \mid Cavity)\ \mathbf{P}(Cavity)}{\sum_{c \in \{\mathrm{t},\mathrm{f}\}} \mathbf{P}(Toothache, Catch \mid Cavity = c) P(Cavity = c)}$$

Probabilistic modelling and inference: summary

- Full joint probability distribution + definition of conditional probability + sum rule allow to compute **any** probability...
  ...but the effort for estimating the distribution and **computational complexity** of inference are too high
- Absolute independence (rare) and conditional independence (common) reduce effort and computational complexity
- Conditional independence can be exploited by considering causal knowledge $P(\mathit{effects}|\mathit{cause})$ (cause–effect relations)
- Diagnostic inference can be carried out from causal knowledge by means of Bayes' rule