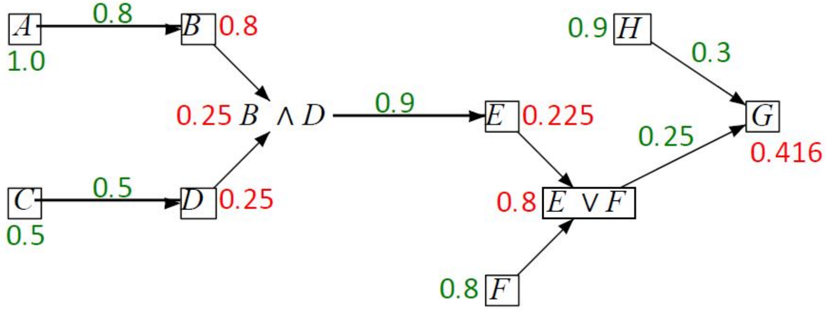
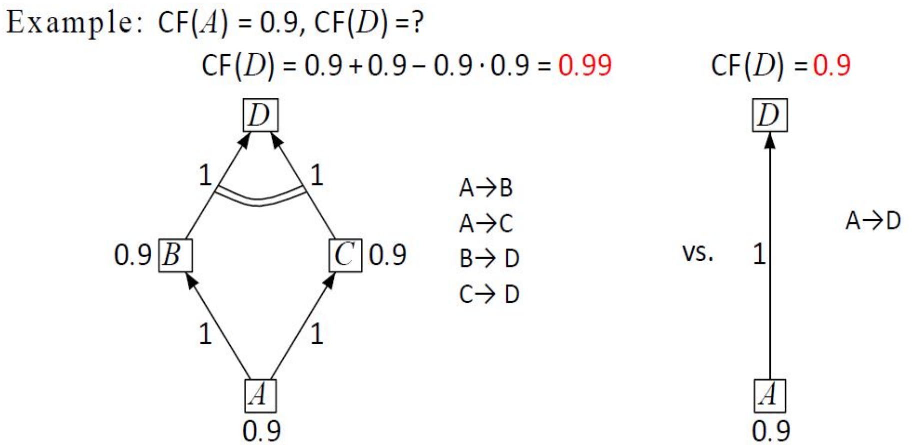
Basics of Applied Probability Theory

Problems with CF-Factors





- Meaning of the CF –Numbers and operators semantically unclear
- Abduction not included
  Given uncertain A $\rightarrow$ B and B. What is the updated knowledge about A?
- Incorrect handling of dependences, see message passing example above
- Is classical probability theory a solution? How to solve the complexity problems?

  Mycin: 500 attributes with more that two values. How to handle 2^{500} probabilities?

Universe of Discourse

We conduct an experiment that has a set E of possible outcomes, e.g.

- Rolling a dice ($E = \{1, 2, 3, 4, 5, 6\}$)
- Arrivals of phone calls ($E = \mathbb{N}_0$)
- Bread roll weights ($E = \mathbb{R}_+$)

An outcome is called an **elementary event**.

All possible elementary events are called the **frame of discernment** or **universe of discourse** E.

The set representation stresses the following facts:

- All possible outcomes are covered by the elements of E. (collectively exhaustive).
- Every possible outcome is represented by exactly one element of E. (mutual disjoint).

Events

Often, we are interested in ***higher-level*** events
(e. g. casting an odd number, arrival of at least 5 phone calls or purchasing a bread roll heavier than 80 grams)

Any subset $A \subseteq E$ is called an event which occurs, if the outcome of the random experiment lies in A.

Since events are sets, we can define for two events A and B:

- $A \cup B$ occurs if A or B occurs; $A \cap B$ occurs if A and B occurs.
- A occurs if A does not occur (i.e., if $\Omega \setminus A$ occurs).
- A and B are ***mutually exclusive***, iff $A \cap B = \emptyset$.

Event Algebra

A (finite) family of sets $\mathcal{E} = \{A_1, \ldots, A_n\}$ is called an **event algebra** on E, if the following conditions hold:

- The certain event E lies in $\mathcal{E}$.
- If $A \in \mathcal{E}$, then $\overline{A} = E \setminus A \in \mathcal{E}$.
- If A_1 and A_2 is in $\mathcal{E}$, then $A_1 \cup A_2 \in \mathcal{E}$ and $A_1 \cap A_2 \in \mathcal{E}$

We can think of (E,$\mathcal{E}$) as a **measurement space**, in which we operate with the outcomes of an experiment. In this course the number of outcomes is normally finite, so we use the power set of E as an appropriate event algebra.

Remark
In the case of a non-finite number of outcomes the concept of a sigma algebra ist used, for example the Borel-algebra for real-valued outcomes.

Probability Function

Given an finite event algebra, we would like to assign to every event $A \in \mathcal{E}$ its **probability** (frequency, subjective belief, …)

A **probability function** $P : \mathcal{E} \rightarrow [0, 1]$ is a mapping that satisfies the so-called Kolmogorov Axioms:

- $\forall A \in \mathcal{E} : 0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- For pairwise disjoint events A, B in $\mathcal{E}$ the equality

  $P(A \cup B) = P(A) + P(B)$ holds.

Remark: In the general case of a sigma algebra we have to replace the third condition (additivity) by the canonically extended concept of sigma-additivity.

Random Variable

For handling Bayes Networks it is more convenient to separate the modelling of the „randomness" from the modelling of the „measurement".

We model
- the randomness by using a probability Q on a set Ω,
- the measurement space E, equipped with the event algebra $\mathcal{E}$,
- the whole random experiment by a **random variable.**

A **random variable X** is a mapping from the probability space Ω to the measurement space E,

$$X: \Omega \rightarrow E$$

If Q is a probability on Ω, then it induces a probability P for all $A \in \mathcal{E}$ by

$$P(A) = Q(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

We are mainly interested in the elementary probabilities $P(\{e\})$ for $e \in E$. Often it is not necessary to give a concrete meaning to the ω's, but it is extremely important to use this intuition.
In the course we use the following short notions (if the context is clear):

$$P(x) = P(X = x) = P(\{X = x\}) = P(\{\omega \in \Omega \mid X(\omega) = x\}), \; x \in E$$

Example Rolling a dice twice

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

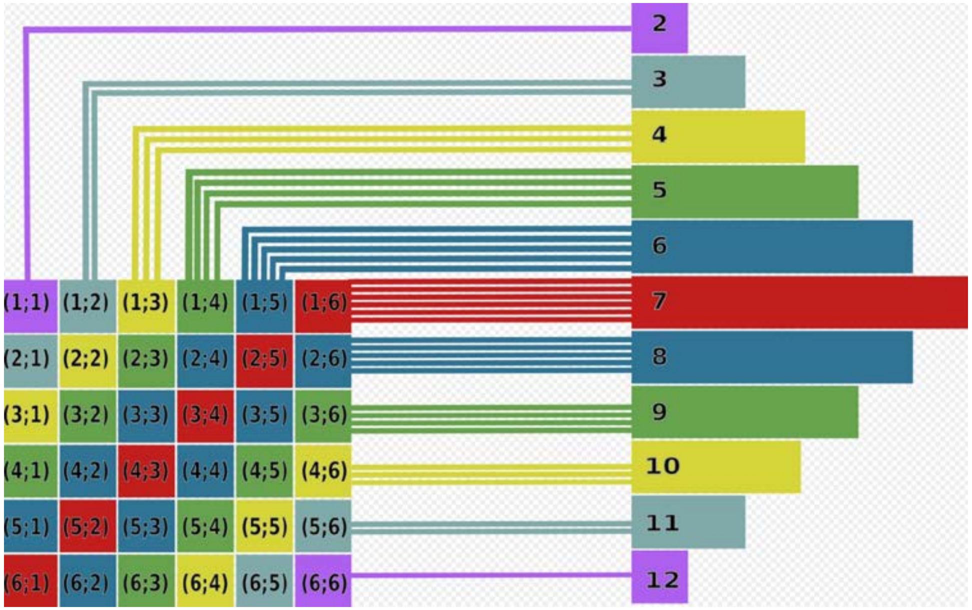
$E = \{2,3,4,\ldots,12\}$

$Q(i,j) = 1/36$, fair dice

$X_1(i,j) = i$, first roll

$X_2(i,j) = j$, second roll

$X(i,j) = i+j$



$P(3) = Q(\{\omega \in \Omega \mid X(\omega)=3\}=Q(\{ (1,2),(2,1)\}=2/36$

$P(1) = 0$, $P(2)=1/36$,.. ,$P(7)=6/36$,…, $P(12)=1/36$

Why (Kolmogorov) Axioms?

If P models an *objectively* observable probability, these axioms are obviously reasonable.

However, why should an agent obey formal axioms when modeling degrees of (subjective) belief?

Objective vs. subjective probabilities

Axioms constrain the set of beliefs an agent can abide.

Ramsey (1926) gave a plausible argument why subjective beliefs should respect Kolmogorov axioms:

The so called "Dutch Book Arguments"

Unconditional Probabilities

$P(A)$ designates the *unconditioned* or *a priori* probability that $A \subseteq \Omega$ occurs if *no* other additional information is present.
For example:

$$P(\mathsf{cavity}) = 0.1$$

A formally different way to state the same would be via a binary random variable Cavity:

$$P(\mathsf{Cavity} = \mathsf{true}) = 0.1$$

A priori probabilities are derived from statistical surveys or general rules.

Unconditional Probabilities

In general a random variable can assume more than two values:

$$
\begin{aligned}
P(\ \text{Weather} = \text{sunny}\ ) &= 0.7 \\
P(\ \text{Weather} = \text{rainy}\ ) &= 0.2 \\
P(\ \text{Weather} = \text{cloudy}) &= 0.02 \\
P(\ \text{Weather} = \text{snowy}\ ) &= 0.08 \\
P(\text{Headache} = \text{true}\ ) &= 0.1
\end{aligned}
$$

$P(X)$ designates the vector of probabilities for the (ordered) domain of the random variable X:

$$
\begin{aligned}
P(\text{Weather}) &= \langle 0.7, 0.2, 0.02, 0.08 \rangle \\
P(\text{Headache}) &= \langle 0.1, 0.9 \rangle
\end{aligned}
$$

Both vectors define the respective probability distributions of the two random variables.

Conditional Probabilities

New evidence can alter the probability of an event.

Example: The probability for cavity increases if information about a toothache arises.

With additional information present, the a priori knowledge must not be used!

$P(A \mid B)$ designates the *conditional* or *a posteriori* probability of A *given* the sole observation (*evidence*) B.

$$P(\mathsf{cavity} \mid \mathsf{toothache}) = 0.8$$

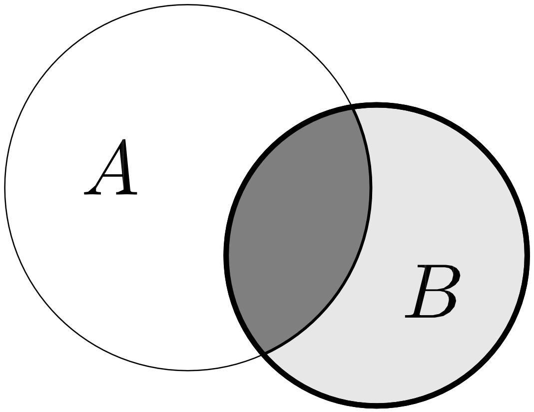
For random variables X and Y $P(X \mid Y)$ represents the set of conditional distributions for each possible value of Y.

Conditional Probabilities

$P(\text{Weather} \mid \text{Headache})$ consists of the following table:

	$\text{h} \mathrel{\widehat{=}} \text{Headache} = \text{true}$	$\neg\text{h} \mathrel{\widehat{=}} \text{Headache} = \text{false}$
$\text{Weather} = \text{sunny}$	$P(\text{W} = \text{sunny} \mid \text{h})$	$P(\text{W} = \text{sunny} \mid \neg\text{h})$
$\text{Weather} = \text{rainy}$	$P(\text{W} = \text{rainy} \mid \text{h})$	$P(\text{W} = \text{rainy} \mid \neg\text{h})$
$\text{Weather} = \text{cloudy}$	$P(\text{W} = \text{cloudy} \mid \text{h})$	$P(\text{W} = \text{cloudy} \mid \neg\text{h})$
$\text{Weather} = \text{snowy}$	$P(\text{W} = \text{snowy} \mid \text{h})$	$P(\text{W} = \text{snowy} \mid \neg\text{h})$

Note that we are dealing with *two* distributions now!
Therefore each column sums up to unity!



A and B are subsets of E

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

Product Rule: $P(A \wedge B) = P(A \mid B) \cdot P(B)$

Also: $P(A \wedge B) = P(B \mid A) \cdot P(A)$

A and B are called ***independent*** iff

$$P(A \mid B) = P(A) \text{ and } P(B \mid A) = P(B)$$

Equivalently, iff the following equation holds true:

$$P(A \wedge B) = P(A) \cdot P(B)$$

Caution: P (A | B) = 0.8 does not mean, that P (A) = 0.8, given B holds

Joint Probabilities

Let $X_1, \ldots, X_n$ be random variables over the same frame of discernment Ω and event algebra $\mathcal{E}$. Then $\vec{X} = (X_1, \ldots, X_n)$ is called a *random vector* with

$$\vec{X}(\omega) \; = \; (X_1(\omega), \ldots, X_n(\omega))$$

Shorthand notation:

$$P(\vec{X} = (x_1, \ldots, x_n)) \; = \; P(X_1 = x_1, \ldots, X_n = x_n) \; = \; P(x_1, \ldots, x_n)$$

Definition:

$$\begin{aligned} P(X_1 = x_1, \ldots, X_n = x_n) \; &= \; P\Big(\Big\{ \omega \in \Omega \;\Big|\; \bigwedge_{i=1}^{n} X_i(\omega) = x_i \Big\}\Big) \\ &= \; P\Big(\bigcap_{i=1}^{n} \{X_i = x_i\} \Big) \end{aligned}$$

Joint Probabilities

Example: $P(\mathsf{Headache}, \mathsf{Weather})$ is the *joint probability distribution* of both random variables and consists of the following table:

	h $\widehat{=}$ Headache = true	¬h $\widehat{=}$ Headache = false
Weather = sunny	$P(\mathsf{W} = \mathsf{sunny} \wedge \mathsf{h})$	$P(\mathsf{W} = \mathsf{sunny} \wedge \neg\mathsf{h})$
Weather = rainy	$P(\mathsf{W} = \mathsf{rainy} \wedge \mathsf{h})$	$P(\mathsf{W} = \mathsf{rainy} \wedge \neg\mathsf{h})$
Weather = cloudy	$P(\mathsf{W} = \mathsf{cloudy} \wedge \mathsf{h})$	$P(\mathsf{W} = \mathsf{cloudy} \wedge \neg\mathsf{h})$
Weather = snowy	$P(\mathsf{W} = \mathsf{snowy} \wedge \mathsf{h})$	$P(\mathsf{W} = \mathsf{snowy} \wedge \neg\mathsf{h})$

All table cells sum up to unity.

Calculating with Joint Probabilities

In the finite case all desired probabilities can be computed from a joint probability distribution.

	toothache	¬toothache
cavity	0.04	0.06
¬cavity	0.01	0.89

Example:
$$\begin{aligned} P(\text{cavity} \lor \text{toothache}) &= P(\text{cavity} \land \text{toothache}) \\ &+ P(\neg\text{cavity} \land \text{toothache}) \\ &+ P(\text{cavity} \land \neg\text{toothache}) = 0.11 \end{aligned}$$

Marginalizations:
$$\begin{aligned} \mathrm{P}(\text{cavity}) &= P(\text{cavity} \land \text{toothache}) \\ &+ P(\text{cavity} \land \neg\text{toothache}) = 0.10 \end{aligned}$$

Conditioning:
$$P(\text{cavity} \mid \text{toothache}) = \frac{P(\text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.04}{0.04 + 0.01} = 0.80$$

Problems

Easiness of computing all desired probabilities comes at an unaffordable price:

Given n random variables with k possible values each, the joint probability distribution contains k^n entries which is infeasible in practical applications.

Hard to handle.

Hard to estimate.

Therefore:

1. Is there a more ***dense*** representation of joint probability distributions?

2. Is there a more ***efficient*** way of processing this representation?

   The answer is *no* for the general case, however, certain dependencies and independencies can be exploited to reduce the number of parameters to a practical size. This is the case in many real world problems.

Two events A and B are called *stochastically independent* iff

$$P(A \wedge B) = P(A) \cdot P(B)$$

$$\Leftrightarrow$$

$$P(A \mid B) = P(A) = P(A \mid \overline{B})$$

Two random variables X and Y are *stochastically independent* iff

$$\forall x \in \mathrm{dom}(X) : \forall y \in \mathrm{dom}(Y) : \quad P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

$$\Leftrightarrow$$

$$\forall x \in \mathrm{dom}(X) : \forall y \in \mathrm{dom}(Y) : \quad P(X = x \mid Y = y) = P(X = x)$$

Shorthand notation: $P(X, Y) = P(X) \cdot P(Y)$.

Note the difference: $P(A) \in [0, 1]$ whereas $P(X) \in [0, 1]^{|\mathrm{dom}(X)|}$.

Conditional Independence

Let X, Y and Z be three random variables. We call X *and* Y *conditionally independent given* Z, iff the following condition holds:

$$\forall x \in \mathrm{dom}(X) : \forall y \in \mathrm{dom}(Y) : \forall z \in \mathrm{dom}(Z) :$$
$$P(X = x, Y = y \mid Z = z) \;=\; P(X = x \mid Z = z) \cdot P(Y = y \mid Z = z)$$

Shorthand notation: $X \perp\!\!\!\perp_P Y \mid Z$

Let $\boldsymbol{X} = \{A_1, \ldots, A_k\}$, $\boldsymbol{Y} = \{B_1, \ldots, B_l\}$ and $\boldsymbol{Z} = \{C_1, \ldots, C_m\}$ be three disjoint sets of random variables. We call $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *conditionally independent given* $\boldsymbol{Z}$, iff

$$P(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z}) \;=\; P(\boldsymbol{X} \mid \boldsymbol{Z}) \cdot P(\boldsymbol{Y} \mid \boldsymbol{Z}) \Leftrightarrow P(\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{Z}) \;=\; P(\boldsymbol{X} \mid \boldsymbol{Z})$$

Shorthand notation: $\boldsymbol{X} \perp\!\!\!\perp_P \boldsymbol{Y} \mid \boldsymbol{Z}$

Conditional Independence

The complete condition for $\boldsymbol{X} \perp\!\!\!\perp_P \boldsymbol{Y} \mid \boldsymbol{Z}$ would read as follows:

$$
\begin{aligned}
&\forall a_1 \in \text{dom}(A_1) : \cdots \forall a_k \in \text{dom}(A_k) : \\
&\quad \forall b_1 \in \text{dom}(B_1) : \cdots \forall b_l \in \text{dom}(B_l) : \\
&\quad\quad \forall c_1 \in \text{dom}(C_1) : \cdots \forall c_m \in \text{dom}(C_m) : \\
&\quad\quad\quad P(A_1 = a_1, \ldots, A_k = a_k, B_1 = b_1, \ldots, B_l = b_l \mid C_1 = c_1, \ldots, C_m = c_m) \\
&\quad\quad\quad\quad = P(A_1 = a_1, \ldots, A_k = a_k \mid C_1 = c_1, \ldots, C_m = c_m) \\
&\quad\quad\quad\quad\; \cdot P(B_1 = b_1, \ldots, B_l = b_l \mid C_1 = c_1, \ldots, C_m = c_m)
\end{aligned}
$$

Remarks:

1. If $\boldsymbol{Z} = \emptyset$ we get (unconditional) independence.
2. We do not use curly braces ($\{\}$) for the sets if the context is clear. Likewise, we use X instead of $\boldsymbol{X}$ to denote sets.

Example Conditional Independence

- $\mathrm{dom}(G) = \{\mathsf{mal}, \mathsf{fem}\}$ Geschlecht (gender)
- $\mathrm{dom}(S) = \{\mathsf{sm}, \overline{\mathsf{sm}}\}$ Raucher (smoker)
- $\mathrm{dom}(M) = \{\mathsf{mar}, \overline{\mathsf{mar}}\}$ Verheiratet (married)
- $\mathrm{dom}(P) = \{\mathsf{preg}, \overline{\mathsf{preg}}\}$ Schwanger (pregnant)

p_{GSMP}		G = mal		G = fem	
		$S = \mathsf{sm}$	$S = \overline{\mathsf{sm}}$	$S = \mathsf{sm}$	$S = \overline{\mathsf{sm}}$
$M = \mathsf{mar}$	$P = \mathsf{preg}$	0	0	0.01	0.05
	$P = \overline{\mathsf{preg}}$	0.04	0.16	0.02	0.12
$M = \overline{\mathsf{mar}}$	$P = \mathsf{preg}$	0	0	0.01	0.01
	$P = \overline{\mathsf{preg}}$	0.10	0.20	0.07	0.21

Example Conditional Independence

$$P(\mathsf{G}=\mathsf{fem}) = P(\mathsf{G}=\mathsf{mal}) = 0.5 \qquad P(\mathsf{P}=\mathsf{preg}) = 0.08$$

$$P(\mathsf{S}=\mathsf{sm}) = 0.25 \qquad P(\mathsf{M}=\mathsf{mar}) = 0.4$$

Gender and Smoker are not independent:

$$P(\mathsf{G}=\mathsf{fem} \mid \mathsf{S}=\mathsf{sm}) = 0.44 \neq 0.5 = P(\mathsf{G}=\mathsf{fem})$$

Gender and Marriage are marginally independent but conditionally dependent given Pregnancy:

$$P(\mathsf{fem}, \mathsf{mar} \mid \overline{\mathsf{preg}}) \approx 0.152 \quad \neq \quad 0.169 \approx P(\mathsf{fem} \mid \overline{\mathsf{preg}}) \cdot P(\mathsf{mar} \mid \overline{\mathsf{preg}})$$

Bayes Theorem

Product Rule (for events A and B):

$$P(A \cap B) = P(A \mid B)P(B) \quad \text{and} \quad P(A \cap B) = P(B \mid A)P(A)$$

Equating the right-hand sides:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

For random variables X and Y:

$$\forall x \forall y : \quad P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y)P(Y = y)}{P(X = x)}$$

Generalization concerning background knowledge/evidence E:

$$P(Y \mid X, E) = \frac{P(X \mid Y, E)P(Y \mid E)}{P(X \mid E)}$$

Example : Bayes Theorem

P (toothache | cavity) = 0.4

P (cavity) = 0.1

P(toothache) = 0.05

$$P(\text{cavity} \mid \text{toothache}) = \frac{0.4 \cdot 0.1}{0.05} = 0.8$$

Why not estimate P (cavity | toothache) right from the start?

Causal knowledge like P (toothache | cavity) is more robust than diagnostic knowledge P (cavity | toothache).

The causality P (toothache | cavity) is independent of the a priori probabilities P(toothache) and P(cavity).

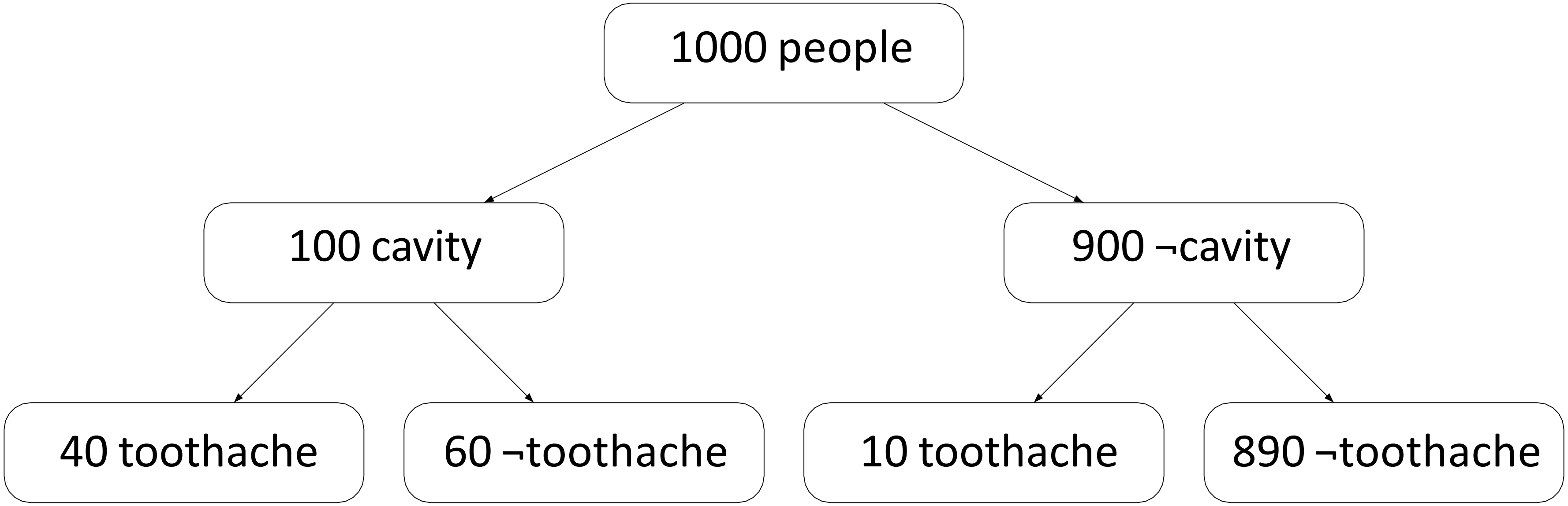
If P (cavity) rose in a caries epidemic, the causality P (toothache | cavity) would remain constant whereas both P (cavity | toothache) and P (toothache) would increase according to P(cavity).

A physician, after having estimated P (cavity | toothache), would not know a rule for updating.

Example : Using absolute Numbers

$P\,(\text{toothache} \mid \text{cavity}) = 0.4 \qquad P\,(\text{cavity}) = 0.1$

$P\,(\text{toothache} \mid \neg\text{cavity}) = \frac{1}{90} \qquad P\,(\text{cavity} \mid \text{toothache}) = \frac{40}{40+10} = 0.8$



$$P\,(c \mid t) = \frac{P\,(t \mid c) \cdot P\,(c)}{P(t)} = \frac{P\,(t \mid c) \cdot P\,(c)}{P(t \mid c) \cdot P(c) + P(\text{t} \mid \neg c) \cdot P(\neg c)}$$

Remark: ¬ means not

Example 3: Relative Probabilities

Assumption:
We would like to consider the probability of the diagnosis GumDisease as well.

$$P\,(\text{toothache} \mid \text{gumdisease}) = 0.7$$
$$P\,(\text{gumdisease}) = 0.02$$

Which diagnosis is more probable?

If we are interested in *relative probabilities* only (which may be sufficient for some decisions), P (toothache) needs not to be estimated:

$$\frac{P\,(c \mid t)}{P(g \mid t)} = \frac{P\,(t \mid c)P\,(c)}{P\,(t)} \cdot \frac{P\,(t)}{P\,(t \mid g)P\,(g)}$$

$$= \frac{P\,(t \mid c)P\,(c)}{P\,(t \mid g)\,P\,(g)} = \frac{0.4 \cdot 0.1}{0.7 \cdot 0.02}$$

$$= 28.57$$

Example 3: Normalization

If we are interested in the absolute probability of $P(c \mid t)$ but do not know $P(t)$, we may conduct a complete case analysis (according c) and exploit the fact that $P(c \mid t) + P(\neg c \mid t) = 1$.

$$P(c \mid t) = \frac{P(t \mid c) P(c)}{P(t)}$$

$$P(\neg c \mid t) = \frac{P(t \mid \neg c) P(\neg c)}{P(t)}$$

$$1 = P(c \mid t) + P(\neg c \mid t) = \frac{P(t \mid c) P(c)}{P(t)} + \frac{P(t \mid \neg c) P(\neg c)}{P(t)}$$

$$P(t) = P(t \mid c) P(c) + P(t \mid \neg c) P(\neg c)$$

Example 3: Normalization

Plugging into the equation for $P(c \mid t)$ yields

$$P(c \mid t) = \frac{P(t \mid c)\, P(c)}{P(t \mid c)\, P(c) + P(t \mid \neg c)\, P(\neg c)}$$

For general random variables, the equation read:

$$P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y)\, P(Y = y)}{\sum_{\forall z \in \mathrm{dom}(Y)} P(X = x \mid Y = z)\, P(Y = z)}$$

Example 3: Multiple Evidences

The patient complains about a toothache. From this first evidence the dentist infers:

$$P(\text{cavity} \mid \text{toothache}) = 0.8$$

The dentist palpates the tooth with a metal probe which catches into a fracture:

$$P(\text{cavity} \mid \text{fracture}) = 0.95$$

Both conclusions might be inferred via Bayes rule. But what does the combined evidence yield? Using Bayes rule further, the dentist might want to determine

$$P(\text{cavity} \mid \text{toothache} \wedge \text{fracture}) = \frac{P(\text{toothache} \wedge \text{fracture} \mid \text{cavity}) \cdot P(\text{cavity})}{P(\text{toothache} \wedge \text{fracture})}$$

Problem:
He needs P(toothache∧catch | cavity), i.e. diagnostics knowledge for all combinations of symptoms in general. Better is to incorporate evidencesstep-by-step:

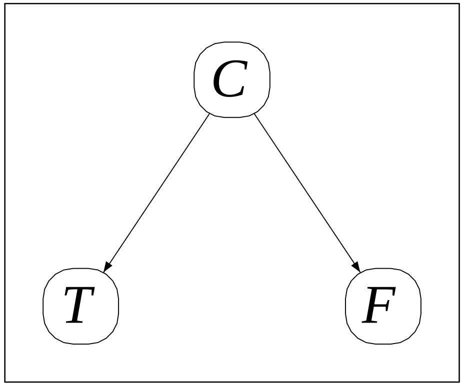
$$P(Y \mid X, E) = \frac{P(X \mid Y, E)P(Y \mid E)}{P(X \mid E)}$$

Abbreviations

C — Cavity

T — Toothache

F — Fracture



Objective:
Computing $P(C \mid T, F)$ with just using information about $P(\cdot \mid C)$ and under exploitation of independence relations among the variables.

Note that capital letter indicates that C is a random variable, in the example with values c and ¬c

Example 3: Multiple Evidences

A priori: $P(c)$

Evidence toothache: $P(c \mid t) = P(c) \dfrac{P(t \mid c)}{P(t)}$

Evidence fracture: $P(c \mid t, f) = P(c \mid t) \dfrac{P(f \mid c, t)}{P(f \mid t)}$

Information about conditional independence

$$P(f \mid c, t) = P(f \mid c)$$

$$P(c \mid t, f) = P(c) \frac{P(t \mid c)}{P(t)} \frac{P(f \mid c)}{P(f \mid t)}$$

Seems that we still have to cope with symptom interdependencies?!

Example 3: Multiple Evidences

Compound equation from last slide:

$$P(c \mid t, f) = P(c)\,\frac{P(t \mid c)\,P(f \mid c)}{P(t)\,P(f \mid t)}$$

$$= P(c)\,\frac{P(t \mid c)\,P(f \mid c)}{P(f, t)}$$

$P(f, t)$ is a normalizing constant and can be computed if $P(f \mid \neg c)$ and $P(t \mid \neg c)$ are known

$$P(f, t) = P(f, t \mid c)\,P(c) + P(f, t \mid \neg c)\,P(\neg c)$$

Therefore, using conditional independence, we finally arrive at the following solution...

Example 3: Multiple Evidences

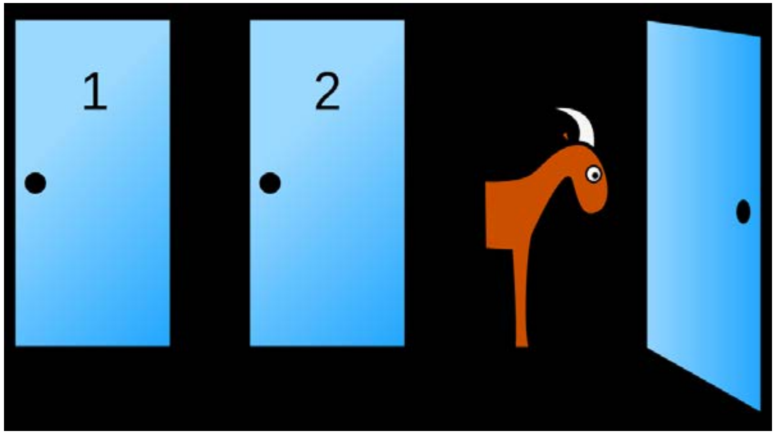
$$P(c \mid f, t) = \frac{P(c) \quad P(t \mid c) \quad P(f \mid c)}{P(f \mid c)\ P(t \mid c)\ P(c) + P(f \mid \neg c)\ P(t \mid \neg c)\ P(\neg c)}$$

Note that we only use probabilities $P(\cdot \mid c)$ together with the a priori (marginal) probabilities $P(c)$ and $P(\neg c)$.

Example 4:Monty Hall Problem

Marylin Vos Savant in her riddle column in the New York Times:

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?



Example 4: Monty Hall Puzzle

The question in this form is under-determined; the correct answer depends on what additional assumptions are made.

Vos Savant replied, “Yes, you should switch. The goal chosen first has a 1⁄3 chance of winning, but the second goal has a 2⁄3 chance of winning. ”

Vos Savant's answer is correct, but only under an additional assumption: Regardless of whether the Porsche or a goat is behind the gate initially chosen by the candidate, the showmaster must open a non-chosen gate with a goat and offer the change.

Even with this additional assumption, it is counter-intuitive for many people that the chance of winning actually increases to 2⁄3 instead of just 1⁄2. As a result, according to their own estimates, vos Savant received ten thousand letters, most of which doubted the correctness of their answer.

Example 5: Simpson's Paradox

Example: c = Patient takes medication, e = patient recovers,

	e	$\neg e$	Σ	Recovery rate
c	20	20	40	50%
$\neg c$	16	24	40	40%
	36	44	80	

Men	e	$\neg e$	Σ	Rec.rate	Women	e	$\neg e$	Σ	Rec.rate
c	18	12	30	60%	c	2	8	10	20%
$\neg c$	7	3	10	70%	$\neg c$	9	21	30	30%
	25	15	40			11	29	40	

$$P(e \mid c) > P(e \mid \neg c)$$

$$P(e \mid c, m) < P(e \mid \neg c, m)$$

$$P(e \mid c, w) < P(e \mid \neg c, w)$$

Note: P is an estimated probability based on the relative fequencies of relevant patients

Focusing vs. Revision

Philosophical topics, studied already by Kant, Gärdenfors

Example for **Focusing/Conditioning**

- Prior knowledge: a fair dice
- New evidence: the result is an odd number
- A posteriori knowledge via focusing: conditional probability

Underlying probability space is not changed

Example for **Belief Change/Revision**

- Prior knowledge: a fair dice
- New evidence: weight near the 6
- Belief change via revision

Underlying probability space is modified

Excursus: Causality vs. Correlation

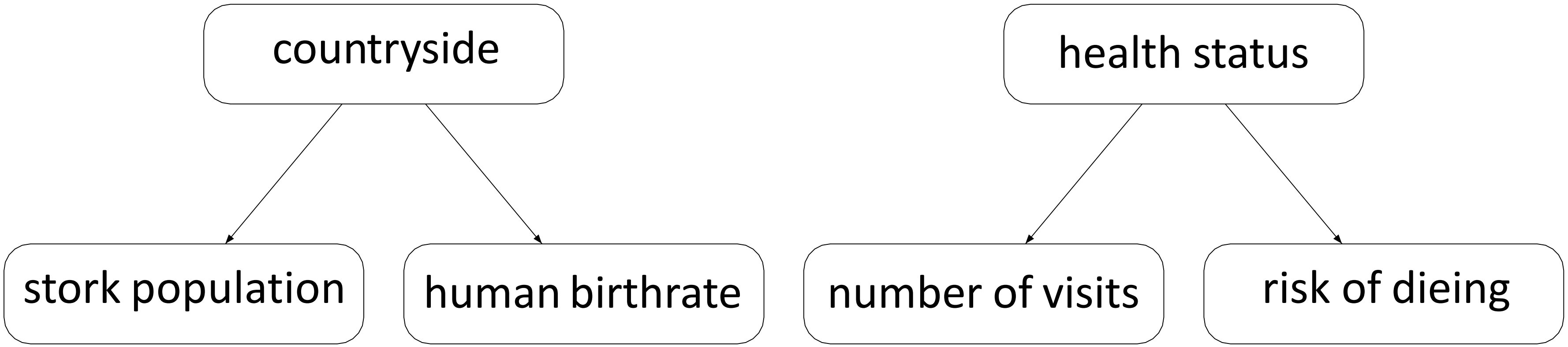
Another old philosophical topic, studied e.g. by Aristoteles, still under discussion

Press acceleration pedal → car is faster (causality)

Stork population high → human birthrate (correlation, but no causality)

Visit doctor often → high risk of dieing (correlation, but no causality)

„Causal" Explanation of the correlations is possible by adding attributes:



Probabilistic Reasoning is complex

Probabilistic reasoning is difficult and may be problematic:

- Probabilistic methods are not **truth functional**,

  $i.e.$ $P(a \wedge b)$ is not determined by $P(a)$ and $P(b)$

  Example: From $P(a) = P(b) = 0.5$ we can i.g. only conclude $P(a \wedge b) \in [0, 0.5]$

- Central problem in real applications: How does **additional information** affect the current knowledge? I.e., if $P(B \mid A)$ is known, what can be said about $P(B \mid A \wedge C)$?
- High complexity: n propositions → 2^n full conjunctives

  Hard to specify these probabilities.

- Probabilistic Reasoning in high dimensions is complex with

  lots of pitfalls.

Summary

Uncertainty is inevitable in complex and dynamic scenarios that force agents to cope with ignorance.

Probabilities express the agent's inability to vote for a definitive decision. They use the degree of belief.

If an agent violates the axioms of probability, it may exhibit irrational behavior in certain circumstances.

The Bayes rule is used to derive unknown probabilities from present knowledge and new evidence.

Multiple evidences can be effectively included into computations by exploiting conditional independencies.