TDT4171 Artificial Intelligence Methods

Lecture 2 – Uncertainty

Norwegian University of Science and Technology

Helge Langseth
Gamle Fysikk 255
helge.langseth@ntnu.no

NTNU

Outline

1. Uncertainty
   - Why consider uncertainty?
   - Probability
   - Syntax and Semantics
   - Inference
   - Independence and Bayes' Rule
2. Summary

Announcements

- **Exercise-set 1 is out:**
  - First of ten assignments; seven needed.
  - Delivery using **using Blackboard**.
  - Deadline **Thursday 30/01 at 23:59**.
  - To pass: "Decent attempt" on everything.
  - Collaboration not allowed.
  - Studass availability: Check Blackboard. In-person and Teams
  - **Problems?** First student assistants, then TAs (tdt4171@idi.ntnu.no).
- **RefGrp:** Is established! Names and email of members on BB (menu choice "Course information")

- Strong AI — Weak AI
- Chinese room – A “falsification” of Strong AI (??)
- The Turing Test
- It is actually useful to think about these things!

From the announcement of Norway’s National AI strategy (2020)

Chapter 12: Uncertainty

- Main message of the Chapter: *"Uncertainty is everywhere"*
- Treatment here differs from the basic statistics course
  - We consider **high-dimensional** distributions
  - **More** focus on efficient representation & inference
  - **More** focus on modelling
  - **More** focus on making decisions
  - **Less** focus on statistical method (hypothesis tests, confidence intervals, etc.)

Basic statistics – The "Sum" and "Product" rules

- **Magic bag:**
  - I have a bag with elements described by `colour` and `shape`, namely `(blue, ball)`, `(red, ball)`, `(blue, box)`, `(red, box)`, and `(yellow, box)`.
  - I draw one item randomly with uniform probability.
  - What is the probability of getting something `blue`?
  - What is the probability of getting something `blue` **or** a `ball`?

- **Throwing dice:**
  1. Throwing a dice once, what is the probability of getting a "6"? **How are you thinking** to get to a quantification?
  2. What is the expected number of dots you get?
  3. Throwing twice, what is prob. of getting "6" **both times**?
  4. Throwing twice, what is prob. of getting "6" **at least once**?
  5. What is the expected number of dots you get throwing **twice**?
  6. Throwing the dice n, times what is the probability of getting "6" **at least once**?

Discuss with your "neighbour" for a couple of minutes.

MAGIC BAG:

	something blue	Blue or ball
(blue, ball)	✓	✓
(red, ball)	–	✓
(blue, box)	✓	✓
(red, box)	–	–
(yellow, box)	–	–
	2 of 5	3 of 5

2 of 5 means $p = \frac{2}{5} = .4$

$p = \frac{3}{5} = 0.6$

DICE

1) 1 out of 6 $= \frac{1}{6}$

2) $E(X) = \sum_x x \cdot P(X=x) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \ldots + \frac{1}{6} \cdot 6 = \underline{\underline{3.5}}$

3) $P(X_1=6 \wedge X_2=6) = P(X_1=6) \cdot P(X_2=6) = \frac{1}{6} \cdot \frac{1}{6} = \underline{\underline{\frac{1}{36}}}$

4) $P(X_1=6 \vee X_2=6) = P(X_1=6) + P(X_2=6) - P(X_1=6 \wedge X_2=6)$
$= 1/6 + 1/6 - 1/36 = \underline{\underline{11/36}}$

5) $E(X_1+X_2) = E(X_1) + E(X_2) = 3.5 + 3.5 = \underline{\underline{7}}$

6) $P(X_1=6 \vee \ldots \vee X_n=6) = 1 - P(X_1 \neq 6 \wedge \ldots \wedge X_n \neq 6)$
$= 1 - (P(X \neq 6))^n = 1 - (5/6)^n$

Uncertainty

Let action A_t = leave for airport t minutes before flight

Interesting question: Will A_t get me there on time?

Problems:

1. Partial observability (road state, other drivers' plans, etc.)
2. Noisy sensors (traffic report on radio)
3. Uncertainty in action outcomes (flat tire, etc.)
4. Immense complexity of modelling and predicting traffic

Uncertainty

Let action A_t = leave for airport t minutes before flight

Interesting question: Will A_t get me there on time?

Consequences for a purely logical approach:

1. Risks falsehood: *"A_{30} will get me there on time"*
2. Leads to conclusions that are too weak for decision making: *"A_{30} will get me there on time if there's no accidents, no queues, it doesn't rain, no police around, etc."*
3. Is uninteresting (A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport...)

Probability

Probabilistic assertions **summarize** effects of

Laziness: Failure to enumerate exceptions, qualifications, etc.

Ignorance: Lack of relevant facts, initial conditions, etc.

Subjective or *Bayesian* probability:

- Probabilities relate propositions to one's own state of knowledge, e.g., $P(A_{30}|\text{no reported accidents}) = 0.7$
- These are **not** claims of a "probabilistic tendency" in the current situation (no "repeated experiments", but might still be learned from past experience of similar situations).
- Probabilities of propositions change with new evidence, e.g., $P(A_{30}|\text{no reported accidents, 5 a.m.}) = 0.85$

Probability – Interpretation

Interpretation of probability – Rationality

If a rational agent believes $X = \texttt{true}$ with probability p, then it would be indifferent to the bet

- Pay p to take part
- If $X = \texttt{true}$, then have a return of 1

$\Rightarrow$ We take the **subjectivist's** stance towards probability in this course. The probabilities measure belief, and are used for making decisions.

Intuition:

$$\begin{aligned}
\mathbb{E}[\text{Earnings}] &= \mathbb{E}[\text{Return} - \text{Bet}] = \mathbb{E}[\text{Return}] - \mathbb{E}[\text{Bet}] \\
&= \underbrace{\{1 \cdot P(\text{win}) + 0 \cdot P(\text{loose})\}}_{\mathbb{E}[\text{Return}]} - \underbrace{p}_{\mathbb{E}[\text{Bet}]} \\
&= P(X = \texttt{true}) - p = 0 \qquad\qquad \Rightarrow \text{Indifference}
\end{aligned}$$

Probability – Interpretation (cont'd)

Interpretation of odds from a bookie

- A bookmaker assumes the probabilities for Home, Draw and Away to be p_H^*, p_D^*, and p_A^*, respectively.
- "Fair" odds would be $\omega_H^* = 1/p_H^*$, $\omega_D^* = 1/p_D^*$, $\omega_A^* = 1/p_A^*$.
- A profit margin (typically $\approx 5\%$) is enforced, so that the given odds are $\omega_H = \alpha \cdot \omega_H^*$, $\omega_D = \alpha \cdot \omega_D^*$, $\omega_A = \alpha \cdot \omega_A^*$.
- We can recover the value of α by using $\alpha = (1/\omega_H + 1/\omega_D + 1/\omega_A)^{-1}$; profit margin is now $1 - \alpha$.

Example: Bournemouth vs. Liverpool , Feb 1st 2025 – Unibet

The given odds are $\boldsymbol{\omega} = (4.50, 4.20, 1.68)$, which means that

$$\alpha = (1/4.50 + 1/4.20 + 1/1.68)^{-1} = 0.947$$

$\Rightarrow$ Unibet expect 5.3% gain **IF** their probabilities are correct.

$\Rightarrow$ Only gamble if you think Unibet's probs are wrong, e.g., $p_H > 1/\omega_H = .22$, when Unibet believes $p_H^* = \alpha/\omega_H = .21$.

Making decisions under uncertainty

Suppose I believe the following:

$$
\begin{aligned}
P(A_{15} \text{ gets me there on time} | \ldots) &= 0.04 \\
P(A_{30} \text{ gets me there on time} | \ldots) &= 0.80 \\
P(A_{75} \text{ gets me there on time} | \ldots) &= 0.99 \\
P(A_{1440} \text{ gets me there on time} | \ldots) &= 0.9999
\end{aligned}
$$

Which action should I choose?

If I am going to Oslo to meet my sister (standard stuff)?

Making decisions under uncertainty

Suppose I believe the following:

$$\begin{aligned} P(A_{15} \text{ gets me there on time}|\ldots) &= 0.04 \\ P(A_{30} \text{ gets me there on time}|\ldots) &= 0.80 \\ P(A_{75} \text{ gets me there on time}|\ldots) &= 0.99 \\ P(A_{1440} \text{ gets me there on time}|\ldots) &= 0.9999 \end{aligned}$$

Which action should I choose?

If I am going to Oslo to meet my sister (standard stuff)?

If I am going to Oslo for a major event (say, a concert)?

Making decisions under uncertainty

Suppose I believe the following:

$$
\begin{aligned}
P(A_{15} \text{ gets me there on time}|\ldots) &= 0.04 \\
P(A_{30} \text{ gets me there on time}|\ldots) &= 0.80 \\
P(A_{75} \text{ gets me there on time}|\ldots) &= 0.99 \\
P(A_{1440} \text{ gets me there on time}|\ldots) &= 0.9999
\end{aligned}
$$

Which action should I choose?

Best answer – **for me** – depends on **my** preferences for missing my flight (missing appointment) vs. waiting (boring).

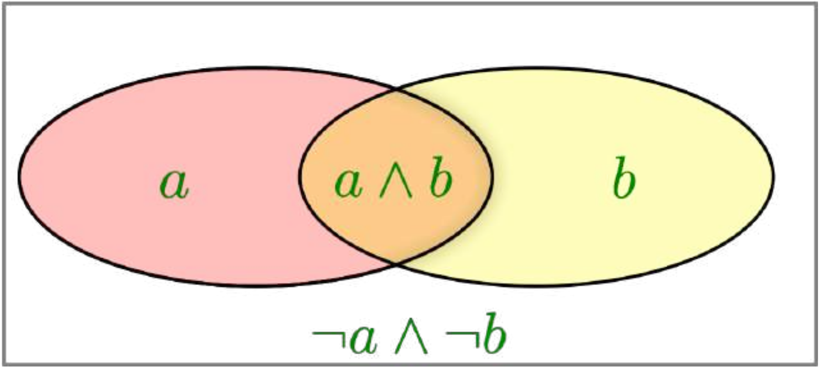
Utility theory: Represent and infer preferences.

Decision theory: Combo of utility theory and probability theory.

Plan: Look at probability theory for some weeks, then utility theory afterwards.

Why use probability calculus?

The definitions imply that certain logically related events must have related probabilities, e.g., $P(a \lor b) = P(a) + P(b) - P(a \land b)$.



de Finetti (1931): An agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money **regardless of outcome**.

de Finetti: Example

Assume an agent will throw two dice, and has the following beliefs:

- Belief in a single die being 6 is $1/6$
- Belief in at least one die being 6 is $1/2$ (**should be** $11/36$ **...**)

The beliefs makes agent willing to accept these simultaneous bets:

1. Pay $5/6$, get 1 if first die is not 6
2. Pay $5/6$, get 1 if second die is not 6
3. Pay $1/2$, get 1 if at least one die is 6

What happens?

de Finetti: Example

Assume an agent will throw two dice, and has the following beliefs:

- Belief in a single die being 6 is $1/6$
- Belief in at least one die being 6 is $1/2$ (**should be** $11/36 \dots$)

The beliefs makes agent willing to accept these simultaneous bets:

1. Pay $5/6$, get 1 if first die is not 6
2. Pay $5/6$, get 1 if second die is not 6
3. Pay $1/2$, get 1 if at least one die is 6

What happens?

No. times 6	Pays	Return	Earnings
Zero	$13/6$	2	-1/6
Once	$13/6$	2	-1/6
Twice	$13/6$	1	-7/6

Syntax for propositions

Propositional or Boolean random variables, e.g.,

- $\mathtt{Cavity}$ (do I have a cavity?)
- $\mathtt{Cavity} = \mathtt{true}$ is a proposition, also written $\mathtt{cavity}$

Discrete random variables (finite or infinite)

- e.g., $\mathtt{Weather}$ is one of $\langle \mathtt{rain}, \mathtt{sunny}, \mathtt{cloudy}, \mathtt{snow} \rangle$
- $\mathtt{Weather} = \mathtt{sunny}$ is a proposition
- Values must be exhaustive and mutually exclusive

Continuous random variables (bounded or unbounded)

- e.g., $\mathtt{Temp} = 11.6$; also allow, e.g., $\mathtt{Temp} < 12.0$.

Prior probability

Prior or unconditional probabilities of propositions

e.g., $P(\texttt{Cavity} = \texttt{true}) = 0.2$ and $P(\texttt{Weather} = \texttt{rain}) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

$\mathbf{P}(\texttt{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalized, i.e., sums to 1)

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

$\mathbf{P}(\texttt{Weather}, \texttt{Cavity})$ is now a 4×2 matrix of values:

$\texttt{Weather} =$	$\texttt{rain}$	$\texttt{sunny}$	$\texttt{cloudy}$	$\texttt{snow}$	Sum
$\texttt{Cavity} = \texttt{true}$	0.144	0.02	0.016	0.02	0.2
$\texttt{Cavity} = \texttt{false}$	0.576	0.08	0.064	0.08	0.8
Sum	0.720	0.10	0.080	0.10	1.0

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Recap from basic statistics – Conditional distributions

Card-trick: I have three cards: One blue on both sides, one orange on both sides, and one with one blue and one orange side. I am going to pick a card at random (uniform probability), and place one of the sides down (again chosen at random, each side equally probable).

1. What is the probability of the colour of the side facing down being orange?
2. You get to know the side facing up is orange. What is the probability of the colour of the side facing down being orange **now**?

Discuss with your neighbour for a couple of minutes.

CARD TRICK

	SIDE 1	2
THREE CARDS: CARD 1	B	B
CARD 2	O	O
CARD 3	B	O

1) SIX "POSSIBLE WORLDS": EITHER FACE OF EITHER CARD.

3 ORANGE CARD-SIDES → $P = 3/6 = 1/2$

2) SIDE UP IS ORANGE. POSSIBLE WORLDS:

* CARD 2 IS CHOSEN, SIDE 1 UP → O DOWN
* CARD 2 IS CHOSEN, SIDE 2 UP → O DOWN
* CARD 3 IS CHOSEN, SIDE 2 UP → B DOWN

→ 2 OF 3 POSSIBILITIES GIVE O DOWN

→ $P = 2/3$.

Conditional probability

Conditional or posterior probabilities, e.g.,

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

means "80% *probability of* `cavity` *given that* `toothache` *is all I know*", **NOT** "*if* `toothache` *(and maybe some other information) then* 80% *chance of* `cavity`"

Notation for conditional distributions:

$\mathbf{P}(\texttt{Cavity} \mid \texttt{Toothache}) = 2$-element vector of 2-element vectors

Conditional probability

Conditional or posterior probabilities, e.g.,

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

means "80% *probability of* `cavity` *given that* `toothache` *is all I know*", **NOT** "*if* `toothache` *(and maybe some other information) then* 80% *chance of* `cavity`"

Notation for conditional distributions:

$\mathbf{P}(\texttt{Cavity} \mid \texttt{Toothache}) = 2$-element vector of 2-element vectors

Note:

The less specific belief remains valid after more evidence arrives, but is not always useful.

If we know more, e.g., `cavity` is also given, then we have

$$P(\texttt{cavity} \mid \texttt{toothache}, \texttt{cavity}) = 1$$

Conditional probability

Conditional or posterior probabilities, e.g.,

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

means "80% *probability of* `cavity` *given that* `toothache` *is all I know*", **NOT** "*if* `toothache` *(and maybe some other information) then* 80% *chance of* `cavity`"

Notation for conditional distributions:

$\mathbf{P}(\texttt{Cavity} \mid \texttt{Toothache})$ = 2-element vector of 2-element vectors

Note also:

New evidence may be irrelevant, allowing simplification, e.g.,

$$P(\texttt{cavity} \mid \texttt{toothache}, \textbf{winRBK}) = P(\texttt{cavity} \mid \texttt{toothache}) = 0.8$$

This kind of inference, sanctioned by domain knowledge, is crucial

Conditional probability

Definition of conditional probability:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

Product rule gives an alternative formulation:

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

A general version holds for whole distributions, e.g.,

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

(View as a 4×2 set of equations, **not** matrix multiplication)

Conditional probability

Definition of conditional probability:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

Product rule gives an alternative formulation:

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

A general version holds for whole distributions, e.g.,

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

(View as a 4×2 set of equations, **not** matrix multiplication)

The **Chain rule** is derived by successive application of product rule:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1, \ldots, X_{n-1})\, \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots, X_{n-2})\, \mathbf{P}(X_{n-1}|X_1, \ldots, X_{n-2})\, \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \ldots = \prod_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1})\end{aligned}$$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

Inference by enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

$$P(\texttt{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

$$P(\texttt{cavity} \vee \texttt{toothache}) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

Inference by enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

Can also compute conditional probabilities:

$$\begin{aligned} P(\neg\texttt{cavity} \mid \texttt{toothache}) &= \frac{P(\neg\texttt{cavity} \wedge \texttt{toothache})}{P(\texttt{toothache})} \\ &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} \\ &= 0.4 \end{aligned}$$

Recap from statistics – Monty Hall

- You're given the choice between Doors a, b, and c in a game-show. There is a price behind only one door. You select Door a, but are not allowed to open it.
- The host is to open one door, observing the following rules:
  - He cannot open the door you have chosen
  - He cannot open the door in front of the price.

1. What is the joint probability table over the two variables `PrizeLocation` and `HostOpens` given that you selected a?
2. The host decides to open door b. What is the probability distribution $\mathbf{P}(\texttt{PrizeLocation} \mid \texttt{HostOpens} = \texttt{b})$?

Team up with your neighbour for a couple of minutes.

MONTY HALL

DOORS: A, B, C. YOU SELECTED A

JOINT PROBABILITY

		HOST OPENS			SUM
		A	B	C	
LOCATION	A	0	1/6	1/6	1/3
	B	0	0	1/3	1/3
	C	0	1/3	0	1/3
	SUM	0	1/2	1/2	1

$$P(\text{Location} \mid \text{Opens}=b) = \frac{P(\text{Location}, \text{opens}=b)}{P(\text{opens}=b)}$$

$$= [1/6, 0, 1/3] / \tfrac{1}{2} = \underline{\underline{\left[\tfrac{1}{3}, 0, \tfrac{2}{3}\right]}}$$

Normalization

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Denominator can be viewed as a normalization constant α:

$$\mathbf{P}(\texttt{Catch} \mid \texttt{toothache}) = \alpha\, \mathbf{P}(\texttt{Catch}, \texttt{toothache})$$

$$= \alpha\, [\mathbf{P}(\texttt{Catch}, \texttt{toothache}, \texttt{cavity})$$

$$+\mathbf{P}(\texttt{Catch}, \texttt{toothache}, \neg\texttt{cavity})]$$

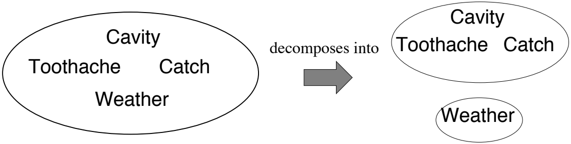
$$= \alpha\, [\langle 0.108 + 0.016, 0.012 + 0.064 \rangle]$$

$$= \alpha\, \langle 0.124, 0.076 \rangle = \langle 0.62, 0.38 \rangle$$

General idea: Compute distribution on query variable by fixing evidence variables and summing over hidden variables

Independence

A and B are **independent** iff
$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or
$\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



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

32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$

Absolute independence **extremely powerful but extremely rare**.

Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

Conditional independence

$\mathbf{P}(\mathtt{Toothache}, \mathtt{Cavity}, \mathtt{Catch})$ has $2^3 - 1 = 7$ independent entries

1. If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
   $P(\mathtt{catch}|\mathtt{toothache}, \mathtt{cavity}) = P(\mathtt{catch}|\mathtt{cavity})$
2. The same independence holds if I haven't got a cavity:
   $P(\mathtt{catch} \mid \mathtt{toothache}, \neg\mathtt{cavity}) = P(\mathtt{catch}|\neg\mathtt{cavity})$

Catch is conditionally independent of Toothache given Cavity:

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

Equivalent statements:

1. $\mathbf{P}(\mathtt{Toothache}|\mathtt{Catch}, \mathtt{Cavity}) = \mathbf{P}(\mathtt{Toothache}|\mathtt{Cavity})$
2. $\mathbf{P}(\mathtt{Toothache}, \mathtt{Catch}|\mathtt{Cavity}) =$
   $\mathbf{P}(\mathtt{Toothache}|\mathtt{Cavity})\mathbf{P}(\mathtt{Catch}|\mathtt{Cavity})$

Conditional independence contd.

Write out full joint distribution using chain rule:

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

We now have $2+2+1=5$ independent numbers

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n.

Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes' Rule

The product rule $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$, can be used to prove Bayes' rule:

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

$$\mathbf{P}(\boldsymbol{Y}|x) = \frac{\mathbf{P}(x|\boldsymbol{Y})\mathbf{P}(\boldsymbol{Y})}{\mathbf{P}(x)} = \alpha \cdot \mathbf{P}(x|\boldsymbol{Y})\mathbf{P}(\boldsymbol{Y})$$

Bayes' rule is useful for assessing "diagnostic probability" from "causal probability":

$$\begin{aligned} \mathbf{P}(\mathtt{Cause}|\mathtt{effect}) &= \frac{\mathbf{P}(\mathtt{effect}|\mathtt{Cause})\mathbf{P}(\mathtt{Cause})}{P(\mathtt{effect})} \\ &= \alpha \cdot \mathbf{P}(\mathtt{effect}|\mathtt{Cause}) \cdot \mathbf{P}(\mathtt{Cause}) \end{aligned}$$

Basic statistics – Bayes' rule

- Two factories, **Factory A** and **Factory B**, make light-bulbs. **Factory A** produces 60% of the bulbs, **Factory B** the rest.
- The probability of a light-bulb from **Factory A** being defect is 0.01, from **Factory B** the probability is 0.02.
- What is the probability of a lightbulb being from **Factory A** **given that it is defect**?

Discuss with your neighbour for a couple of minutes.

BAYES RULE

FACTORY A: 60% OF BULBS. DEFECT-PROB 0.01

FACTORY B: 40% OF BULBS. DEFECT-PROB 0.02

$$P(\text{Factory}=a \mid \text{defect}) = \frac{P(\text{defect} \mid \text{Factory}=a) \cdot P(\text{Factory}=a)}{P(\text{defect})}$$

$$P(\text{defect}) = P(\text{defect} \wedge \text{Factory}=a) + P(\text{defect} \wedge \text{Factory}=b)$$

$$= 0.6 \cdot 0.01 + 0.4 \cdot 0.02 = 0.014$$

$$P(\text{Factory}=a \mid \text{defect}) = \frac{0.01 \cdot 0.6}{0.014} = \underline{0.429}$$

Bayes' Rule and conditional independence

$$\begin{aligned}
&\mathbf{P}(\texttt{Cavity}|\texttt{toothache} \wedge \texttt{catch}) \\
&\quad = \alpha\,\mathbf{P}(\texttt{toothache} \wedge \texttt{catch}|\texttt{Cavity})\mathbf{P}(\texttt{Cavity}) \\
&\quad = \alpha\,\mathbf{P}(\texttt{toothache}|\texttt{Cavity})\mathbf{P}(\texttt{catch}|\texttt{Cavity})\mathbf{P}(\texttt{Cavity})
\end{aligned}$$

This is an example of a so-called **naïve Bayes** model:

$$\mathbf{P}(\texttt{Cause}, \texttt{Effect}_1, \ldots, \texttt{Effect}_n) = \mathbf{P}(\texttt{Cause}) \prod_i \mathbf{P}(\texttt{Effect}_i|\texttt{Cause})$$

Total number of parameters is linear in no. effects!!!

Summary of Chapter 12

- **Probability** is a rigorous formalism for uncertain knowledge
- **Joint probability distribution** specifies probability of every **atomic event**
- Queries can be answered by **summing** over atomic events
- For nontrivial domains, we must find a way to **reduce** the size of the joint distribution, as it grows like $O(d^n)$
- **Independence** and **conditional independence** provide the tools for simplification.
- Calculations can be rather heavy; next we will **start using a SW tool**, which does this for us