What is probabilistic conditioning?

Good Afternoon, AI@RU
Dec. 6, 2018

Bart Jacobs — Radboud University
bart@cs.ru.nl

Outline

Introduction

States and predicates

Inference via channels in Bayesian networks

Constructive and destructive updating

A sketch of the quantum case

Conclusions

Where we are, so far

Introduction

States and predicates

Inference via channels in Bayesian networks

Constructive and destructive updating

A sketch of the quantum case

Conclusions

Huh? What? Who? Here?

Bien étonnés de se trouver ensemble

I'll briefly explain

Own background

- student of Mathematics & Philosopy; PhD in theoretical computer science (1991), with Henk Barendregt
  - background in logic, type theory, category theory
- subsequent research in Java program semantics and verification
  - partly parallel theoretical work on state-based systems in terms of "coalgebras"
  - applications to Java-based smart cards
- this led to much work in (applied) security & privacy
  - with lots of public exposure
- ERC Advanced Grant (2013): Quantum Computation, Logic & Security
  - much emphasis on quantum, but also classical, probability
  - leading to new axiomatisation: effectus theory
  - also connections to quantum cognition

Boosting my credentials in SocSci

☺ B. Jacobs, Quantum Effect Logic in Cognition, *Journal of Mathematical Psychology* 81, 2017, p.1–10.

Two entangled points that I wish to make

(1) **Updating (belief revision) is very subtle**
- quantum updating differs from classical probabilistic updating
  - e.g. successive quantum updates do not commute
  - there is "lower" and "upper" conditioning, see BJ, QPL'18.
- But also: there are different forms of classical updating: "constructive" and "destructive"
  - destructive updatings also do not commute
  - are they useful in cognition?

(2) **We need a good language & logic for probability**
- standard probabilistic notation $P(-)$ is confusing
- distributions ("states") are usually left implicit
- states and predicates are not distinguished — and hence state and predicate transformation are not recognised

Ad 1: More about updating

- Basic fact: successive quantum updates & conjunctions do not commute
  - see later for details
- Successive human (cognitive) primings also do not commute
  - that is, the human mind is sensitive to the order in which information is presented
  - this observation is one of the motivations for the area of quantum cognition theory
- My favourite example: what impression do you obtain about Bob from the following two orders:

  (1) Alice is pregnant ; Bob visits Alice
  (2) Bob visits Alice ; Alice is pregnant

  Common reaction about Bob: (1) good guy, (2) guilty guy.

Ad 1: More about updating, continued

- Updating happens in the presence of evidence
  - e.g. the test is positive (evidence), so the (a priori) disease probability is updated to the (a posteriori) probability ...
- Evidence can also be soft or probabilistic
  - e.g. I'm 80% sure I heard the alarm
- How to handle (updating with) soft evidence is an unsolved issue in the (classical) probabilistic community
  - it pops up now and then in the literature, in diffent forms
  - two methodologies can be distinguished (Darwiche, Chan):
    (1) "Jeffrey's rule", or also "probability kinematics"
    (2) "Pearl's method for virtual evidence"

    The outcomes are quite different, see later

Ad 1: More about updating, still continued

- A recent new look at this matter in arXiv:1807.05609
  - fully: BJ, *A Mathematical Account of Soft Evidence, and of Jeffrey's 'destructive' versus Pearl's 'constructive' updating*
  - fresh terminology: "destructive" versus "constructive" updating
  - systematic re-description, involving e.g. "daggers" of channels
- Further complication:
  - there is also "intervention" from Pearl's causality work
  - it changes the graph structure ("surgery")
  - not discussed here, but should be included in the larger picture of both updating and intervention, see arXiv:1811.08338
  - the difference is highly relevant in cognition theory, see e.g. Steven Sloman, *Causal Models. How people think about the world and its alternatives*, OUP 2005.

Ad 2. Language for probability

- In this area there are too many fat & sloppy books
  - they calculate like 'headless chickens'
  - precise definitions are often lacking (what is a conjugate prior?)
  - algorithms mostly without specification — let alone verification
- Typically a problem is solved by:
  - writing down a some formula with lot's of $P(-)$'s and $P(- \mid -)$'s
  - calculating the outcome, e.g. via multiplication, summation/integration, normalisation
  - omitting explanation of the methodology
- A high-level account gives more conceptual clarity
  - what are the relevant concepts/notions, like states & predicates
  - what are the basic operations, like state/predicate transformation, and updating
  - how to combine these operations in clearly structured expressions

Ad 2. Language for probability, continued

- Relevant articles towards a systematic language
  - BJ and F. Zanasi, *The Logical Essentials of Bayesian Reasoning*, arXiv:1804.01193
  - BJ, *A Mathematical Account of Soft Evidence, and of Jeffrey's 'destructive' versus Pearl's 'constructive' updating*, arXiv:1807.05609
- One of the main embarrassements of the field is that there is no widely accepted and useful probabilistic symbolic logic
  - with proper syntax and deduction rules
  - with well-defined semantics, preferably working uniformly for discrete, continuous and quantum probability
- As a step towards this goal, a (uniform) library called EfProb now exists in Python, for probabilistic calculations
  - see efprob.cs.ru.nl
  - joint work with former PhD-student Kenta Cho

Ad 2. Language for probability, still continued

Relevant quote from Pearl'89

> *To those trained in traditional logics, symbolic reasoning is the standard, and nonmonotonicity a novelty. To students of probability, on the other hand, it is symbolic reasoning that is novel, not nonmonotonicity. Dealing with new facts that cause probabilities to change abruptly from very high values to very low values is a commonplace phenomenon in almost every probabilistic exercise and, naturally, has attracted special attention among probabilists. The new challenge for probabilists is to find ways of abstracting out the numerical character of high and low probabilities, and cast them in linguistic terms that reflect the natural process of accepting and retracting beliefs.*

Indeed, a symbolic logic with both updating and non-monotonicity is non-standard and non-trivial.

Plan for today

(1) Background info about states and predicates, state- and predicate-transformation, and on (constructive) updating
(2) Example usage of this language for inference in Bayesian networks
(3) Destructive versus constructive updating
(4) A quantum example

Along the way the power of a more abstract language will be demonstrated.

- following Wittenstein's *zeigen* instead of *sagen*

Where we are, so far

Introduction

States and predicates

Inference via channels in Bayesian networks

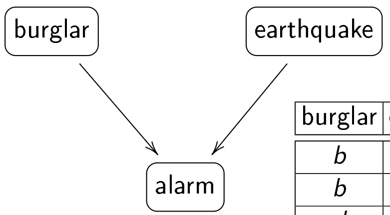
Constructive and destructive updating

A sketch of the quantum case

Conclusions

Example I, from Barber'12

P(burglar)
0.01



P(earthquake)
0.000001

burglar	earthquake	P(alarm)
b	e	0.9999
b	$\sim e$	0.99
$\sim b$	e	0.99
$\sim b$	$\sim e$	0.0001

Barber in his 2012 book asks (in Example 3.1/3.2):

> *Imagine that we are 70% sure we heard the alarm sounding. What is the probability of a burglary?*

- "destructive" computation: 69% (Barber)
- "constructive" computation: 2%

Which one is right?

Discrete probability distributions / states

Notation

- Fair coin: $\frac{1}{2}|H\rangle + \frac{1}{2}|T\rangle$
- Fair dice: $\frac{1}{6}|1\rangle + \frac{1}{6}|2\rangle + \frac{1}{6}|3\rangle + \frac{1}{6}|4\rangle + \frac{1}{6}|5\rangle + \frac{1}{6}|6\rangle$

ket notation

- $|-\rangle$ is pure syntactic sugar — stemming from quantum
- more confusing to omit them, as in: $\frac{1}{6}1 + \frac{1}{6}2 + \frac{1}{6}3 + \frac{1}{6}4 + \frac{1}{6}5 + \frac{1}{6}6$

- Write $\mathcal{D}(X)$ for the set of such probability distributions $\sum_i r_i|x_i\rangle$ where $x_i \in X$, $r_i \in [0,1]$ with $\sum_i r_i = 1$
- Distributions $\omega \in \mathcal{D}(X)$ will often be called states of X

Predicates, as fuzzy functions

- A predicate on a set X is a function $p\colon X \to [0,1]$
  - such predicates will be used as soft evidence, by default
- It is called sharp (non-fuzzy) if $p(x) \in \{0,1\}$ for each $x \in X$
  - sharp predicates are indicator functions $\mathbf{1}_E$ for an "event" $E \subseteq X$
- There are "truth", "falsum", "orthosupplement" predicates
  - e.g. $(p^\perp)(x) = 1 - p(x)$, so that $p^{\perp\perp} = p$
  - then: $(\mathbf{1}_E)^\perp = \mathbf{1}_{\neg E}$
  - the set $[0,1]^X$ of predicates on X forms an effect module
- There is also fuzzy conjunction $p \,\&\, q$ via pointwise multiplication
  - $(p \,\&\, q)(x) = p(x) \cdot q(x)$
  - then $\mathbf{1}_E \,\&\, \mathbf{1}_D = \mathbf{1}_{E\cap D}$
  - this makes $[0,1]^X$ a commutative monoid in the category of effect modules

Products and marginalisations of states

- For states $\omega_1 \in \mathcal{D}(X_1)$ and $\omega_2 \in \mathcal{D}(X_2)$ we can form the product state $\omega_1 \otimes \omega_2 \in \mathcal{D}(X_1 \times X_2)$ by:

$$(\omega_1 \otimes \omega_2)(x_1, x_2) \;=\; \omega_1(x_1) \cdot \omega_2(x_2)$$

- For a joint state $\sigma \in \mathcal{D}(X_1 \otimes X_2)$ there are marginalisations $M_i(\sigma) \in \mathcal{D}(X_i)$, given by:

$$M_1(\sigma)(x_1) \;=\; \sum_{x_2} \sigma(x_1, x_2) \qquad M_2(\sigma)(x_2) \;=\; \sum_{x_1} \sigma(x_1, x_2)$$

- It is to easy that marginalisation after product returns the originals:

$$M_1(\omega_1 \otimes \omega_2) \;=\; \omega_1 \qquad M_2(\omega_1 \otimes \omega_2) \;=\; \omega_2$$

- A joint state is called non-entwined if it is the product of its marginals; most joint states are entwined.

Combining states and predicates

Let $\omega \in \mathcal{D}(X)$ be state/distribution, $p \in [0,1]^X$ a predicate, both on X.

- **Validity** $\omega \models p$, in $[0,1]$
  - defined as $\sum_x \omega(x) \cdot p(x)$
  - also known as expected value of p in state ω
  - Traditional notation: $P(A)$ is $\omega \models \mathbf{1}_A$ for $A \subseteq X$
- **Conditioning** $\omega|_p$, in $\mathcal{D}(X)$
  - assuming validity $\omega \models p$ is non-zero
  - defined as: $\omega|_p = \sum_x \frac{\omega(x) \cdot p(x)}{\omega \models p}|x\rangle$
  - Traditionally $P(B \mid A)$ is $\omega|_{\mathbf{1}_A} \models \mathbf{1}_B$

Validity and conditioning example

- Take $X = \{1,2,3,4,5,6\}$ with state dice $\in \mathcal{D}(X)$
  - recall dice $= \frac{1}{6}|1\rangle + \frac{1}{6}|2\rangle + \frac{1}{6}|3\rangle + \frac{1}{6}|4\rangle + \frac{1}{6}|5\rangle + \frac{1}{6}|6\rangle$
- Take even predicate $\mathbf{1}_E \in [0,1]^X$ for $E = \{2,4,6\} \subseteq X$; it's sharp
  - define odd via: $O = \neg E = \{1,3,,5\}$, so that $\mathbf{1}_O = (\mathbf{1}_E)^\perp$.
- dice $\models \mathbf{1}_E \;=\; \frac{1}{6}\cdot 0 + \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 0 + \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 0 + \frac{1}{6}\cdot 1 \;=\; \frac{1}{2}$
- $\text{dice}|_{\mathbf{1}_E} \;=\; \frac{1/6}{1/2}|2\rangle + \frac{1/6}{1/2}|4\rangle + \frac{1/6}{1/2}|6\rangle = \frac{1}{3}|2\rangle + \frac{1}{3}|4\rangle + \frac{1}{3}|6\rangle$
- $\text{dice}|_{\mathbf{1}_E} \models \mathbf{1}_O \;=\; 0$

Two basic laws of conditioning — for soft evidence

Recall that we write $p \,\&\, q$ for the pointwise product $(p \,\&\, q)(x) = p(x)\cdot q(x)$ of predicates $p, q \in [0,1]^X$.

product rule $\qquad \omega|_p \models q \;=\; \dfrac{\omega \models p \,\&\, q}{\omega \models p}$

Bayes' rule $\qquad \omega|_p \models q \;=\; \dfrac{(\omega|_q \models p)\cdot(\omega \models q)}{\omega \models p}$

Easy but important observation:

These rules are equivalent, using that & is commutative
(the rules differ in a quantum setting)

State and predicate transformation

A channel $X \to Y$ is a function $X \to \mathcal{D}(Y)$

- it captures conditional probability $p(y \mid x)$ as $x \mapsto p(y \mid x)$
- alternatively, it is a stochastic matrix
- For a state $\omega \in \mathcal{D}(X)$ we get $c \gg \omega \in \mathcal{D}(Y)$ via:

$$(c \gg \omega)(y) \quad := \quad \sum_x c(x)(y)\cdot\omega(x).$$

- For a predicate $q \in [0,1]^Y$ we have $c \ll q \in [0,1]^X$ by:

$$(c \ll q)(x) \quad := \quad \sum_y c(x)(y)\cdot q(y).$$

Basic relation

$$\omega \models c \ll q \quad = \quad c \gg \omega \models q.$$

Calculus of channels

Channels can be composed sequentially, and in parallel:

- $(d \bullet c)(x) = d \gg c(x)$
- $(e \otimes f)(x,y) = e(x) \otimes f(y)$

- These $\bullet$ and $\otimes$ interact appropriately — abstractly because $\mathcal{K}\ell(\mathcal{D})$ is a symmetric monoidal category
- They also interact well with state and predicate transformation, eg:

$$(d \bullet c) \gg \omega \;=\; d \gg (c \gg \omega) \qquad \text{and} \;\; (d \bullet c) \ll q \;=\; c \ll (d \ll q)$$

Keeping states and predicates apart

- States and predicates look similar and are often confused
  - each state is a predicate: $\mathcal{D}(X) \subseteq [0,1]^X$
  - but not the other way around: predicates may have infinite support, and their probabilities need not add up to one.
- States and predicates have entirely different algebraic structures
  - states on a set X form a convex set
  - predicates on a set X form an effect module
- State transformation preserves convex sums, and predicate transformation preserves the effect module structure.
- Explicitly, for a channel $c\colon X \to \mathcal{D}(Y)$,
  - $c \gg (-)\colon \mathcal{D}(X) \to \mathcal{D}(Y)$ is a map in $\mathbf{Conv} = \mathcal{EM}(\mathcal{D})$
  - $c \ll (-)\colon [0,1]^Y \to [0,1]^X$ is map in $\mathbf{EMod}$

Conditioning and transformation

Overview table for joint work with Fabio Zanasi:

notation	action	terminology
$\omega\big|_{(c \ll q)}$	first do predicate transformation, then update the state	evidential reasoning, or explanation, or backward inference
$c \gg (\omega\big|_p)$	first update the state, then do state transformation	causal reasoning, or prediction, or forward inference

Where we are, so far

Introduction

States and predicates

Inference via channels in Bayesian networks

Constructive and destructive updating

A sketch of the quantum case

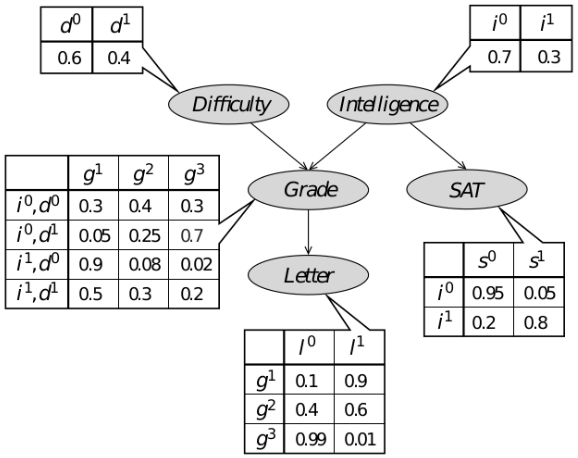
Conclusions

Two main ideas

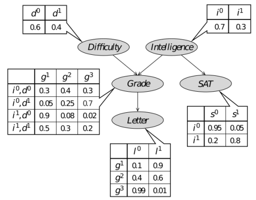
(1) (Brendan Fong) A Bayesian network is a graph of channels
  - formally, in the Kleisli category $\mathcal{K}\ell(\mathcal{D})$ of the distribution monad $\mathcal{D}$
  - or of the Giry monad $\mathcal{G}$ in for continuous probability

(2) (BJ & Fabio Zanasi) Bayesian inference happens via a combination of state/predicate transformation and conditioning
  - using sequential and parallel composition $\bullet$ and $\otimes$ from the Kleisli category

The student example from Koller-Friedman (PGM, 2009)



The student example via channels

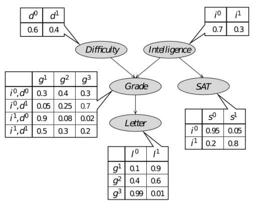


- Use domains $D = \{d^0, d^1\}$, $I = \{i^0, i^1\}$, $G = \{g^0, g^1, g^2\}$, $S = \{s^0, s^1\}$, $L = \{l^0, l^1\}$
- With initial states $\omega_D = 0.6|\, d^0\,\rangle + 0.4|\, d^1\,\rangle$ and $\omega_I = 0.7|\, i^0\,\rangle + 0.3|\, i^1\,\rangle$
- And channels $c_G \colon D \times I \to G$, $c_S \colon I \to S$, $c_L \colon G \to L$

  for instance with:
  $$c_S(i^0) = 0.95|\, s^0\,\rangle + 0.05|\, s^1\,\rangle$$
  $$c_S(i^1) = 0.2|\, s^0\,\rangle + 0.8|\, s^1\,\rangle$$

We discuss some questions from Koller-Friedman from a "transformation & update" perspective — and get the same outcome as in the book, but via systematic procedures/expressions

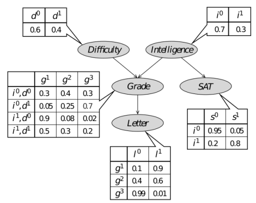
What is the a priory letter probability?



We do forward state transformation:

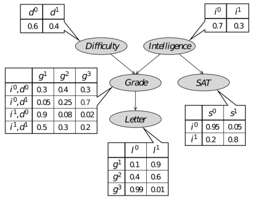
$$\begin{aligned} c_L \gg (c_G \gg (\omega_D \otimes \omega_I)) \\ = 0.498|\, l^0\,\rangle + 0.502|\, l^1\,\rangle \\ = (c_L \bullet c_G) \gg (\omega_D \otimes \omega_I). \end{aligned}$$

What if we know that the student is not intelligent?



- Non-intelligence involves the point / singleton predicate $\mathbf{1}_{\{i^0\}}$ on $I = \{i^0, i^1\}$
  - it is 1 of i^0 and 0 on i^1
- We can use $\mathbf{1}_{\{i^0\}}$ to update the state $\omega_I \in \mathcal{D}(I)$ to $\omega_I|_{\mathbf{1}_{\{i^0\}}}$
- With this we compute as before:
  $$\begin{aligned} c_L \gg (c_G \gg (\omega_D \otimes (\omega_I|_{\mathbf{1}_{\{i^0\}}}))) \\ = 0.611|\, l^0\,\rangle + 0.389|\, l^1\,\rangle \\ = (c_L \bullet c_G) \gg ((\omega_D \otimes \omega_I)|_{\mathbf{1} \otimes \mathbf{1}_{\{i^0\}}}) \end{aligned}$$
- Note that this is forward inference: first update the state, then transform

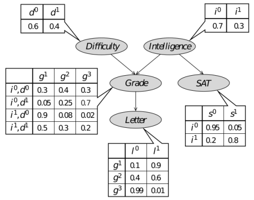
What if we also know that the test is easy?



- We now use the 'easy' predicate $\mathbf{1}_{\{d^0\}}$ on $D = \{d^0, d^1\}$.
- We now also update $\omega_D \in \mathcal{D}(D)$ with this predicate
- Forward inference now gives:

$$\begin{aligned} & c_L \gg \big(c_G \gg \big((\omega_D|_{\mathbf{1}_{\{d^0\}}}) \otimes (\omega_I|_{\mathbf{1}_{\{i^0\}}})\big)\big) \\ & = 0.487|\, l^0\,\rangle + 0.513|\, l^1\,\rangle \\ & = \big(c_L \bullet c_G\big) \gg \big((\omega_D \otimes \omega_I)|_{(\mathbf{1}_{\{d^0\}} \otimes \mathbf{1}_{\{i^0\}})}\big). \end{aligned}$$

What is the intelligence given a C-grade (g^3)?

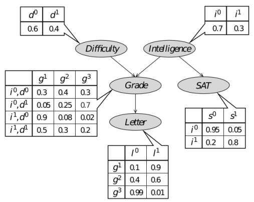


- Evidence predicate is $\mathbf{1}_{\{g^3\}}$ on G
- Predicate transformation along $c_G : D \times I \to G$ gives a predicate $c_G \ll \mathbf{1}_{\{g^3\}}$ on $D \times I$
- We can use it to update $\omega_D \otimes \omega_I$, and then take the second marginal.
- That is:

$$\begin{aligned} & M_2\big((\omega_D \otimes \omega_I)\big|_{c_G \ll \mathbf{1}_{\{g^3\}}}\big) \\ & = 0.921|\, i^0\,\rangle + 0.0789|\, i^1\,\rangle. \end{aligned}$$

- This is backward inference: first transform the predicate, then use it for update

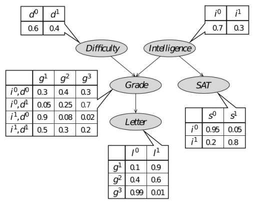
What is the intelligence given a weak recommendation?



- We now start from evidence $\mathbf{1}_{\{l^0\}}$ on L
- We have to do predicate transformation twice to reach the initial states, as in: $c_G \ll (c_L \ll \mathbf{1}_{\{l^0\}})$
- Backward inference now gives:

$$\begin{aligned} & M_2\big((\omega_D \otimes \omega_I)\big|_{c_G \ll (c_L \ll \mathbf{1}_{\{l^0\}})}\big) \\ & = 0.86|\, i^0\,\rangle + 0.14|\, i^1\,\rangle \\ & = M_2\big((\omega_D \otimes \omega_I)\big|_{(c_L \bullet c_G) \ll \mathbf{1}_{\{l^0\}}}\big) \end{aligned}$$

What is the intelligence given a C-grade but a high SAT score?



- We also have evidence $\mathbf{1}_{\{s^0\}}$ on S
- We can transform it to evicence $c_S \ll \mathbf{1}_{\{s^1\}}$ on ω_I, and combine it with the previous evidence on $\omega_D \otimes \omega_I$
- There are several 'logical' ways to do so:

$$\begin{aligned} & M_2\big((\omega_D \otimes (\omega_I|_{c_S \ll \mathbf{1}_{\{s^1\}}}))\big|_{c_G \ll \mathbf{1}_{\{g^3\}}}\big) \\ & = 0.422|\, i^0\,\rangle + 0.578|\, i^1\,\rangle \\ & = M_2\big((\omega_D \otimes \omega_I)\big|_{(\mathbf{1} \otimes (c_S \ll \mathbf{1}_{\{s^1\}})) \& (c_G \ll \mathbf{1}_{\{g^3\}})}\big) \\ & = M_2\big((\omega_D \otimes \omega_I)\big|_{c_G \ll \mathbf{1}_{\{g^3\}}}\big)\big|_{c_S \ll \mathbf{1}_{\{s^1\}}} \end{aligned}$$

Final note on student example

- All inference questions can be answered systematically via "logical" expressions, which can be evaluated
  - in textbooks one usually starts calculating directly
  - more details are in BJ & FZ, arXiv:1804.01193
- The "logical" expressions that we used can also be written in the Python library EfProb
- In fact, a new channel-based inference algorithm has been formulated in this way
  - see arXiv:1804.08032

EfProb code snippets, for student example

Letter given no intelligence:

```
>>> l >> (g >> (d @ (i / ni)))
0.6114|l0> + 0.3886|l1>
```

Intelligence after a C/g3 grade and positive SAT score

```
>>> (d @ (i / (s << ps) )) / (g << cg) % [0,1]
0.4217|i0> + 0.5783|i1>
```


iCIS | Digital Security
Radboud University

Where we are, so far

Introduction

States and predicates

Inference via channels in Bayesian networks

Constructive and destructive updating

A sketch of the quantum case

Conclusions

iCIS | Digital Security
Radboud University

Problem description

- Typical Bayesian inference (reasoning) proceeds as follows:
  - I have "evidence" $E_1, \ldots, E_n$, used to condition my state
  - I then "observe" A, via marginalisation of conditioned state
- The evidence (and observation) are usually "point" or "singleton" predicates
- What if the evidence is "soft"
  - I saw the object in the dark and believe with 30% certainty that it is red and 70% certainty that it is blue
  - How to handle is called soft evidential update problem (Darwiche)
- There are two approaches, giving different outcomes
  - following Jeffrey, renamed as destructive
  - following Pearl, renamed as constructive
    - this is in fact what we have done so far, as $\omega|_p$

iCIS | Digital Security
Radboud University

Example II: Virus – blood pressure

We consider patients having a virus or not, and their blood pressure:

virus?	Low	Medium	High
yes (v)	20%	20%	60%
no ($\sim v$)	60%	30%	10%

We know, as base rate, that 1 in 15 patients have the virus.

Mathematical formalisation:

- underlying domains $V = \{v, \sim v\}$ and $B = \{L, M, H\}$
- prior / base rate distribution $\omega = \frac{1}{15}|\, v\,\rangle + \frac{14}{15}|\sim v\,\rangle$
- channel / Kleisli map $c\colon V \to \mathcal{D}(B)$ extracted from table:

$$c(v) = \tfrac{2}{10}|\,L\,\rangle + \tfrac{2}{10}|\,M\,\rangle + \tfrac{6}{10}|\,H\,\rangle \qquad c(\sim v) = \tfrac{6}{10}|\,L\,\rangle + \tfrac{3}{10}|\,M\,\rangle + \tfrac{1}{10}|\,H\,\rangle$$

Point evidence example

- Suppose we have **high** blood pressure evidence
  - what is the updated virus probability (distribution)?
  - typical Bayes' rule problem
- Channel-based solution, with point predicate $\mathbf{1}_{\{H\}}$ on $B = \{L, M, H\}$

$$\omega|_{c \ll \mathbf{1}_{\{H\}}} = 0.3|\,v\,\rangle + 0.7|\sim v\,\rangle$$

  This 30% probability is higher than the base rate $\frac{1}{15} \sim 6.67\%$

- More abstractly, this involves the dagger channel in opposite direction: going from $c\colon V \to \mathcal{D}(B)$ to:

$$\begin{array}{ccc} B & \xrightarrow{c^{\dagger}_{\omega}} & \mathcal{D}(V) \\ y & \longmapsto & \omega|_{c \ll \mathbf{1}_{\{y\}}} \end{array}$$

Soft evidence example — much like Barber's case

Suppose we have 25% certainty of low blood pressure, 25% of medium 50% of high. What is the updated virus probability?

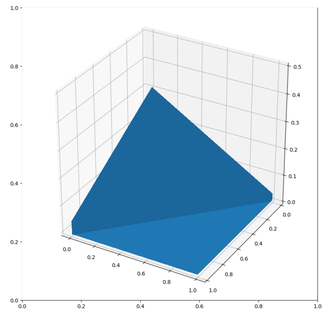
- **Destructive answer**, after Jeffrey
  - Idea: convex combination of point observations
  - $0.25 \cdot$ update with $L + 0.25 \cdot$ update with $M + 0.5 \cdot$ update with H
    $= c^{\dagger}_{\omega} \gg \Big(0.25|\,L\,\rangle + 0.25|\,M\,\rangle + 0.5|\,H\,\rangle\Big)$
    $= 0.0941|\,v\,\rangle + 0.9059|\sim v\,\rangle$
- **Constructive answer**, after Pearl
  - Idea: reason backward with evidence as fuzzy predicate
  - define $p \in [0,1]^B$ as $p(L) = p(M) = 0.25$, $p(H) = 0.5$
  - $\omega|_{c \ll p} = 0.1672|\,v\,\rangle + 0.8328|\sim v\,\rangle$

Substantial difference: 9% versus 17%

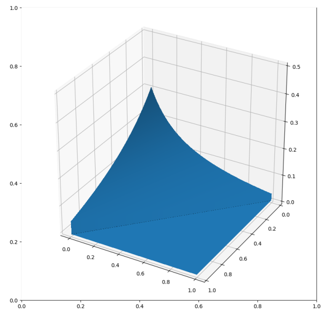
What should decision support systems do — e.g. in medicine?

Plots

We describe the virus probability, given soft evidence $x|\,L\,\rangle + y|\,M\,\rangle + (1-x-y)|\,H\,\rangle$, for $0 \leq x + y \leq 1$ in:



destructive update



constructive update

General observations

Destructive & constructive update coincide on point evidence.

- **Destructive update**
  - interprets soft evidence as state / probability distribution
  - the prior is (largely) overridden by the evidence
  - successive updates do not commute
  - *starting from what you can predict you learn nothing:* $c^{\dagger}_{\omega} \gg (c \gg \omega) = \omega$
- **Constructive update**
  - interprets soft evidence as fuzzy predicate
  - prior is smoothly combined with the evidence — as inner product (following the basic idea: *posterior* $\propto$ *prior* $\cdot$ *likelihood*)
  - successive updates do commute
  - *starting from nothing (constant/uniform predicate) you learn nothing:* $\omega|_{c \ll (r \cdot \mathbf{1})} = \omega$

Constructive and destructive updating

- It is unclear to which form of updating is "the right one"
  - or even what criterion to use
  - let me know if you have ideas and/or more examples
- Intruiging question: which form of updating works best in cognition theory — under which circumstances?

Where we are, so far

Introduction

States and predicates

Inference via channels in Bayesian networks

Constructive and destructive updating

A sketch of the quantum case

Conclusions

Quantum states and predicates

- Let $\mathscr{H}, \mathscr{K}$ be (finite-dimensional) Hilbert spaces
- A state is a density matric $\varrho\colon \mathscr{H} \to \mathscr{H}$
  - this means: $\varrho \geq 0$ and $\mathrm{tr}(\varrho) = 1$
- A predicate is an effect $p\colon \mathscr{H} \to \mathscr{H}$
  - this means: $0 \leq p \leq 1$
  - projections are special "sharp" predicates, with $p^2 = p$
- Orthosupplement is $p^{\perp} = \mathbf{1} - p$, so that $p^{\perp\perp} = p$
- Sequential conjunction $p \mathbin{\&} q := \sqrt{p}\, q \sqrt{p}$ is not commutative
  - projections can be characterised as $p \mathbin{\&} p = p$
  - for projections p, q one gets $p \mathbin{\&} q = p\, q\, p$
- Sequential disjunction $p \mid q := \big(p^{\perp} \mathbin{\&} q^{\perp}\big)^{\perp}$
- Validity is given by Born's rule: $\varrho \models p \;:=\; \mathrm{tr}\big(\varrho\, p\big) \;\in\; [0, 1]$

Linda example (Tverski & Kahneman)

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

What is the likelihood of the following events? Linda is:

(1) active in the feminist movement;

(2) a bank teller;

(3) active in the feminist movement, and a bank teller;

(4) a bank teller, or active in the feminist movement.

The conjunction fallacy concerns the fact that when asked, many people say that option (3) is more likely than option (2), and the disjunction fallacy occurs when option (4) is judged to be less likely than option (1).

Linda example, with a quantum realisation

- Take $\mathscr{H} = \mathbb{C}^2$ with $|v\rangle = \begin{pmatrix} 0.987 \\ -0.1564 \end{pmatrix} \in \mathscr{H}$ giving a state:

$$\omega := |v\rangle\langle v| = \begin{pmatrix} 0.976 & -0.155 \\ -0.155 & 0.024 \end{pmatrix} \in \mathcal{B}(\mathscr{H}).$$

- Use $|u\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|w\rangle = \begin{pmatrix} \cos(2\pi/5) \\ \sin(2\pi/5) \end{pmatrix}$ for predicates

$$\mathrm{fem} := |u\rangle\langle u| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \mathrm{btr} := |w\rangle\langle w| = \begin{pmatrix} 0.095 & 0.293 \\ 0.293 & 0.905 \end{pmatrix}$$

- Then we can describe and compute validities

$$\omega \models \mathrm{fem} = 0.976 \qquad \omega \models \mathrm{btr} = 0.024$$

- The conjunction and disjunction fallacies appear from:

$$\omega \models \mathrm{fem}\ \&\ \mathrm{btr} = 0.09315 \qquad \omega \models \mathrm{btr} \mid \mathrm{fem} = 0.906$$

From: Busemeyer & Bruza, *Quantum Models of Cognition and Decision*, CUP 2012

Where we are, so far

Introduction

States and predicates

Inference via channels in Bayesian networks

Constructive and destructive updating

A sketch of the quantum case

Conclusions

Final remarks

- A symbolic approach has many benefits over direct calculations
  - conceptual clarity
  - better fit with CS/AI, especially for XAI
  - uniformity between different forms of probability (discrete, continuous, quantum)
- Probabilistic updating is a rich area, with many open questions
  - its importance in AI, big data analysis & cognition gives urgency
  - different outcomes for constructive and destructive
  - embarrassingly: explanatory account is missing
- Which approach fits best in cognition theory is unclear
  - can this be established experimentally?
  - possibly a convex combination of constructive and destructive makes sense, depending on the level of certainty
- Many basic research questions in an established area!

Thanks for your attention. Questions/remarks?