Inductive reasoning, conditionals, and belief revision

Part II – The entropy principles and c-representations/c-revisions

Gabriele Kern-Isberner
TU Dortmund, Germany
(partly joint work with Wolfgang Spohn, University of Constance)

Cape-KR 2024, Cape Town, January 2024

Overview of this talk – Part II

- Induction and revision in probabilistics:
  - The entropy principles MaxEnt and MinCEnt

Overview of this talk – Part II

- Induction and revision in probabilistics:
    - The entropy principles MaxEnt and MinCEnt
- Different revision scenarios: conservative revision vs. focusing

Overview of this talk – Part II

- Induction and revision in probabilistics:
  - The entropy principles MaxEnt and MinCEnt
- Different revision scenarios: conservative revision vs. focusing
- Induction and revision with ranking functions:
  - C-representations and c-revisions

Overview of this talk – Part II

- Induction and revision in probabilistics:
  - The entropy principles MaxEnt and MinCEnt
- Different revision scenarios: conservative revision vs. focusing
- Induction and revision with ranking functions:
  - C-representations and c-revisions
- Conclusion and outlook

Overview of this talk – Part II

- Induction and revision in probabilistics:
  - The entropy principles MaxEnt and MinCEnt
- Different revision scenarios: conservative revision vs. focusing
- Induction and revision with ranking functions:
  - C-representations and c-revisions
- Conclusion and outlook

Crucial insights from Part I

$$\Psi \models (B|A) \quad \text{iff} \quad AB \prec_{\Psi} A\overline{B} \quad \text{iff} \quad A \mid\!\sim_{\Psi} B \quad \text{iff} \quad \Psi * A \models B.$$

Crucial insights from Part I

$$\Psi \models (B|A) \quad \text{iff} \quad AB \prec_\Psi A\overline{B} \quad \text{iff} \quad A \mid\!\sim_\Psi B \quad \text{iff} \quad \Psi * A \models B.$$

In a probabilistic environment, we have to adapt these equivalences to take probabilities into account:

- $P \models (B|A)[x]$ iff $P(B|A) = x$ (and $P(A) > 0$) – probabilistic conditionals are interpreted via conditional probabilities.

Crucial insights from Part I

$$\Psi \models (B|A) \quad \text{iff} \quad AB \prec_\Psi A\overline{B} \quad \text{iff} \quad A \mathrel{|\!\sim}_\Psi B \quad \text{iff} \quad \Psi * A \models B.$$

In a probabilistic environment, we have to adapt these equivalences to take probabilities into account:

- $P \models (B|A)[x]$ iff $P(B|A) = x$ (and $P(A) > 0$) – probabilistic conditionals are interpreted via conditional probabilities.
- $A \mathrel{|\!\sim}_P B[x]$ iff $P * A[1] \models B[x]$ iff $P(B|A) = x$ – this amounts to saying that $P * A[1]$ is done via conditioning.

Crucial insights from Part I (cont'd)

Different induction and revision scenarios:

$$(\Psi_\Delta * \mathcal{I}) * \mathcal{I}' = (\mathit{ind}(\Delta) * \mathcal{I}) * \mathcal{I}'$$

- First case: $\mathcal{I}'$ refers to the same context as $\mathcal{I}$, narrowing the context (conservative revision, epistemic expansion). In this case, $\mathcal{I}$ and $\mathcal{I}'$ should be considered on the same level, and we propose $\Psi_\Delta * (\mathcal{I} \cup \mathcal{I}')$.
- Second case: $\mathcal{I}'$ is information on a new, shifted context for which, however, $\mathcal{I}$ is still relevant (update). Then we propose $(\Psi_\Delta * \mathcal{I}) * \mathcal{I}'$ with two revision operators of the same kind.
- Third case: $\mathcal{I}'$ affects background beliefs (learning). If $\mathcal{I}'$ is fully compatible with Δ, we propose $\mathit{ind}(\Delta \cup \mathcal{I}') * \mathcal{I}$, otherwise, we propose $(\mathit{ind}(\Delta) * \mathcal{I}') * \mathcal{I}$.

Crucial insights from Part I (cont'd)

Basing induction on revision via

$$ind = ind_{\Psi_u}$$

Crucial insights from Part I (cont'd)

Basing induction on revision via

$$\mathit{ind} = \mathit{ind}_{\Psi_u} \quad \text{and} \quad \Psi_\Delta = \mathit{ind}(\Delta) = \Psi_u * \Delta.$$

Probabilistic reasoning on maximum entropy

... is an alternative to Bayesian networks which works with weaker assumptions and incomplete probabilistic belief bases $\mathcal{R} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$:

Probabilistic reasoning on maximum entropy

... is an alternative to Bayesian networks which works with weaker assumptions and incomplete probabilistic belief bases $\mathcal{R} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$:

MaxEnt Principle

Maximise indeterminacy (i.e., *entropy*)

$$H(P) = -\sum_{\omega \in \Omega} P(\omega) \cdot \log_2 P(\omega)$$

of a probability distribution P assuming $\mathcal{R}$ as constraints, i.e., solve the optimisation problem

$$(arg) \max_{P \models \mathcal{R}} H(P) = -\sum_{\omega \in \Omega} P(\omega) \cdot \log_2 P(\omega)$$

to obtain the probabilistic model $P^* = ME(\mathcal{R})$ of $\mathcal{R}$ which adds as little information as possible.

MaxEnt inference

MaxEnt inference is a model based (nonmonotonic) inductive inference operator:

C^{ME} inference operator

$$C^{ME}(\mathcal{R}) = \{\phi \in (\mathcal{L} \mid \mathcal{L})^{prob} \mid ME(\mathcal{R}) \models \phi\}$$

C^{ME} satisfies the following properties

- Inclusion/Reflexivity: $\mathcal{R} \subseteq C^{ME}(\mathcal{R})$.
- Cumulativity, i.e., both Cut and Cautious Monotony:

$$\mathcal{R} \subseteq \mathcal{S} \subseteq C^{ME}(\mathcal{R}) \quad \text{implies} \quad C^{ME}(\mathcal{R}) = C^{ME}(\mathcal{S})$$

So, MaxEnt inference satisfies most important axioms of System P.

Probabilistic belief revision via Minimum Cross-Entropy

Remember – nonmonotonic reasoning is closely related to belief revision

Probabilistic belief revision via Minimum Cross-Entropy

Remember – nonmonotonic reasoning is closely related to belief revision

MaxEnt also has a big brother –
use cross-entropy = information distance (= Kullback-Leibler-divergence)

$$R(Q, P) = \sum_{\omega \in \Omega} Q(\omega) \log \frac{Q(\omega)}{P(\omega)}$$

Probabilistic belief revision via Minimum Cross-Entropy

Remember – nonmonotonic reasoning is closely related to belief revision

MaxEnt also has a big brother –
use cross-entropy = information distance (= Kullback-Leibler-divergence)

$$R(Q, P) = \sum_{\omega \in \Omega} Q(\omega) \log \frac{Q(\omega)}{P(\omega)}$$

Belief revision via minimum cross-entropy (MinCEnt)

Given some prior distribution P and some new information $\mathcal{R} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$, choose the unique distribution

$$P^* = P *_{ME} \mathcal{R}$$

that satisfies $\mathcal{R}$ and has minimal information distance to P.

MaxEnt and MinCEnt

The principle of minimum cross-entropy generalizes the principle of maximum entropy: If P_u is a suitable uniform distribution, we have

$$ME(\mathcal{R}) = P_u *_{ME} \mathcal{R}.$$

Hence the ME-methodology is quite a perfect example to illustrate all concepts and relationships presented here in a probabilistic framework.

MaxEnt and MinCEnt

The principle of minimum cross-entropy generalizes the principle of maximum entropy: If P_u is a suitable uniform distribution, we have

$$ME(\mathcal{R}) = P_u *_{ME} \mathcal{R}.$$

Hence the ME-methodology is quite a perfect example to illustrate all concepts and relationships presented here in a probabilistic framework.

The crucial equation for understanding and analyzing ME-revision is

$$P *_{ME} \mathcal{R}(\omega) = \alpha_0 P(\omega) \prod_{\substack{1 \leqslant i \leqslant n \\ \omega \models A_i B_i}} \alpha_i^{1-x_i} \prod_{\substack{1 \leqslant i \leqslant n \\ \omega \models A_i \overline{B_i}}} \alpha_i^{-x_i},$$

where the α_i's (one for each conditional in $\mathcal{R}$) have to be chosen properly to ensure that $P *_{ME} \mathcal{R}$ satisfies all conditionals in $\mathcal{R}$ with the associated probabilities — this is the so-called Success Condition from belief revision.

α_0 is a normalizing factor.

Symbolic reasoning with MaxEnt

Transitive Chaining

$$\frac{\mathcal{R} \ : \ (B|A)[x_1], \ (C|B)[x_2]}{(C|A)[\frac{1}{2}(2x_1x_2 + 1 - x_1)]}$$

Cautious Monotony

$$\frac{\mathcal{R} \ : \ (B|A)[x_1], \ (C|A)[x_2]}{(C|AB)[x_2]}$$

Cut

$$\frac{\mathcal{R} \ : \ (C|AB)[x_1], \ (B|A)[x_2]}{(C|A)[\frac{1}{2}(2x_1x_2 + 1 - x_2)]}$$

Example psychologist

A psychologist summarizes his experiences after working in a helpcenter for addicted people for several years in the following probability distribution:

a : *addicted to alcohol*
d : *addicted to drugs*
y : *being young*

ω	$P(\omega)$	ω	$P(\omega)$	ω	$P(\omega)$	ω	$P(\omega)$
ady	0.050	$\overline{a}dy$	0.333	$ad\overline{y}$	0.053	$\overline{a}d\overline{y}$	0.053
$a\overline{d}y$	0.093	$\overline{a}\overline{d}y$	0.102	$a\overline{d}\overline{y}$	0.225	$\overline{a}\overline{d}\overline{y}$	0.091

Example psychologist (cont'd)

In P, we observe the following conditional probabilities:

$$P(d|a) = 0.242 \quad P(d|\overline{a}) = 0.666$$

Example psychologist (cont'd)

In P, we observe the following conditional probabilities:

$$P(d|a) = 0.242 \quad P(d|\overline{a}) = 0.666$$
$$P(a|y) = 0.246 \quad P(d|y) = 0.662$$

Example psychologist (cont'd)

In P, we observe the following conditional probabilities:

$$P(d|a) = 0.242 \quad P(d|\overline{a}) = 0.666$$
$$P(a|y) = 0.246 \quad P(d|y) = 0.662$$
$$P(a|\overline{y}) = 0.660 \quad P(d|\overline{y}) = 0.251$$

Note that this contains some quantifications of the qualitative belief base $\mathcal{R}_{psycho} = \{(\overline{d}|a), (\overline{a}|d), (\overline{a}|y), (d|y), (\overline{d}|\overline{y}), (a|\overline{y})\}$ from the introduction.

The probabilistic conditionals

$$(d|a)[0.242], (d|\overline{a})[0.666], (a|y)[0.246], (d|y)[0.662]$$

were used to generate P by MaxEnt.

Example psychologist (cont'd)

Now, the psychologist will change his job, he will be working in a clinic in which exclusively people being addicted to alcohol and/or drugs are treated. He knows that in this clinic, the rate of people being addicted to alcohol, but also addicted to drugs is higher than usual and amounts to 40 %.

Example psychologist (cont'd)

Now, the psychologist will change his job, he will be working in a clinic in which exclusively people being addicted to alcohol and/or drugs are treated. He knows that in this clinic, the rate of people being addicted to alcohol, but also addicted to drugs is higher than usual and amounts to 40 %.

Therefore, the psychologist revises his (background) knowledge P with the following new information:

$$\mathcal{R} = \{a \vee d[1], (d|a)[0.4]\}$$

Example psychologist (cont'd)

So he obtains $P^* = P *_{ME} \mathcal{R}$:

ω	$P^*(\omega)$	ω	$P^*(\omega)$	ω	$P^*(\omega)$	ω	$P^*(\omega)$
ady	0.099	$\overline{a}dy$	0.425	$ad\overline{y}$	0.105	$\overline{a}d\overline{y}$	0.066
$a\overline{d}y$	0.089	$\overline{a}\overline{d}y$	0.0	$a\overline{d}\overline{y}$	0.216	$\overline{a}\overline{d}\overline{y}$	0.0

Example psychologist (cont'd)

So he obtains $P^* = P *_{ME} \mathcal{R}$:

ω	$P^*(\omega)$	ω	$P^*(\omega)$	ω	$P^*(\omega)$	ω	$P^*(\omega)$
ady	0.099	$\overline{a}dy$	0.425	$ad\overline{y}$	0.105	$\overline{a}d\overline{y}$	0.066
$a\overline{d}y$	0.089	$\overline{a}\overline{d}y$	0.0	$a\overline{d}\overline{y}$	0.216	$\overline{a}\overline{d}\overline{y}$	0.0

Now we have:

$$P^* \models \quad \begin{array}{ll} (d|a)[0.4] & (d|\overline{a})[1] \\ (a|y)[0.307] & (d|y)[0.855] \end{array}$$

Overview of this talk – Part II

- Induction and revision in probabilistics:
  - The entropy principles MaxEnt and MinCEnt
- Different revision scenarios: conservative revision vs. focusing
- Induction and revision with ranking functions:
  - C-representations and c-revisions
- Conclusion and outlook

Coherence

Coherence Axiom

(Coherence) $\Psi * (\Delta_1 \cup \Delta_2) = (\Psi * \Delta_1) * (\Delta_1 \cup \Delta_2)$.

The axiom of (Coherence)

- demands that adjusting any intermediate epistemic state $\Psi * \Delta_1$ to the full information $\Delta_1 \cup \Delta_2$ should result in the same epistemic state as adjusting Ψ by $\Delta_1 \cup \Delta_2$ in one step.

Coherence

Coherence Axiom

(Coherence) $\Psi * (\Delta_1 \cup \Delta_2) = (\Psi * \Delta_1) * (\Delta_1 \cup \Delta_2)$.

The axiom of (Coherence)

- demands that adjusting any intermediate epistemic state $\Psi * \Delta_1$ to the full information $\Delta_1 \cup \Delta_2$ should result in the same epistemic state as adjusting Ψ by $\Delta_1 \cup \Delta_2$ in one step.
- connects the results of revision based on different prior epistemic states and hence ensures the change process to be coherent.

Coherence

Coherence Axiom

(Coherence) $\Psi * (\Delta_1 \cup \Delta_2) = (\Psi * \Delta_1) * (\Delta_1 \cup \Delta_2)$.

The axiom of (Coherence)

- demands that adjusting any intermediate epistemic state $\Psi * \Delta_1$ to the full information $\Delta_1 \cup \Delta_2$ should result in the same epistemic state as adjusting Ψ by $\Delta_1 \cup \Delta_2$ in one step.
- connects the results of revision based on different prior epistemic states and hence ensures the change process to be coherent.

The axiom has been introduced under the name *Coherence* in [GKI, AIJ 1998] but actually goes back to [Shore & Johnson, 1981].

Coherence

Coherence Axiom

(Coherence) $\Psi * (\Delta_1 \cup \Delta_2) = (\Psi * \Delta_1) * (\Delta_1 \cup \Delta_2)$.

The axiom of (Coherence)

- demands that adjusting any intermediate epistemic state $\Psi * \Delta_1$ to the full information $\Delta_1 \cup \Delta_2$ should result in the same epistemic state as adjusting Ψ by $\Delta_1 \cup \Delta_2$ in one step.
- connects the results of revision based on different prior epistemic states and hence ensures the change process to be coherent.

The axiom has been introduced under the name *Coherence* in [GKI, AIJ 1998] but actually goes back to [Shore & Johnson, 1981].

ME-revision also satisfies (Coherence) [Shore & Johnson, 1981].

Conservative revision and Coherence

Note that (Coherence) does not claim that $(\Psi * \Delta_1) * \Delta_2$ and $(\Psi * \Delta_1) * (\Delta_1 \cup \Delta_2)$ are the same, just to the contrary – the first revised epistemic state is not supposed to maintain prior contextual information, Δ_1, whereas the second should do so, according to success.

Conservative revision and Coherence

Note that (Coherence) does not claim that $(\Psi * \Delta_1) * \Delta_2$ and $(\Psi * \Delta_1) * (\Delta_1 \cup \Delta_2)$ are the same, just to the contrary – the first revised epistemic state is not supposed to maintain prior contextual information, Δ_1, whereas the second should do so, according to success.

However, (Coherence) can help ensuring independence from the representation of background beliefs for inductive reasoning:

Conservative revision and Coherence

Note that (Coherence) does not claim that $(\Psi * \Delta_1) * \Delta_2$ and $(\Psi * \Delta_1) * (\Delta_1 \cup \Delta_2)$ are the same, just to the contrary – the first revised epistemic state is not supposed to maintain prior contextual information, Δ_1, whereas the second should do so, according to success.

However, (Coherence) can help ensuring independence from the representation of background beliefs for inductive reasoning:

Consider the situation where we have

$$\Psi = ind_{\Psi_{bk}}(\Delta) = \Psi_{bk} * \Delta.$$

Conservative revision and Coherence

Note that (Coherence) does not claim that $(\Psi * \Delta_1) * \Delta_2$ and $(\Psi * \Delta_1) * (\Delta_1 \cup \Delta_2)$ are the same, just to the contrary – the first revised epistemic state is not supposed to maintain prior contextual information, Δ_1, whereas the second should do so, according to success.

However, (Coherence) can help ensuring independence from the representation of background beliefs for inductive reasoning:

Consider the situation where we have

$$\Psi = \mathit{ind}_{\Psi_{bk}}(\Delta) = \Psi_{bk} * \Delta.$$

Imagine that we still are aware of the last conditional information Δ_0 that shaped Ψ_{bk}, i.e.,

$$\Psi_{bk} = \Psi_1 * \Delta_0,$$

which would be crucial to know if we want to perform conservative revision $\Psi_1 * (\Delta_0 \cup \Delta)$.

Conservative revision and Coherence (cont'd)

However, in general, $\Psi_{bk} = \Psi_1 * \Delta_0$ and Δ_0 do not determine Ψ_1 uniquely, i.e., there may be a different Ψ_2 satifying also

$$\Psi_{bk} = \Psi_1 * \Delta_0 = \Psi_2 * \Delta_0.$$

Conservative revision and Coherence (cont'd)

However, in general, $\Psi_{bk} = \Psi_1 * \Delta_0$ and Δ_0 do not determine Ψ_1 uniquely, i.e., there may be a different Ψ_2 satifying also

$$\Psi_{bk} = \Psi_1 * \Delta_0 = \Psi_2 * \Delta_0.$$

Hence, also $\Psi_2 * (\Delta_0 \cup \Delta)$ would be a suitable candidate for the outcome of conservative revision.

Conservative revision and Coherence (cont'd)

However, in general, $\Psi_{bk} = \Psi_1 * \Delta_0$ and Δ_0 do not determine Ψ_1 uniquely, i.e., there may be a different Ψ_2 satifying also

$$\Psi_{bk} = \Psi_1 * \Delta_0 = \Psi_2 * \Delta_0.$$

Hence, also $\Psi_2 * (\Delta_0 \cup \Delta)$ would be a suitable candidate for the outcome of conservative revision.

Here (Coherence) guarantees that the resulting epistemic state after revision does not depend on selecting Ψ_1 or Ψ_2:

$$\begin{aligned} \Psi_1 * (\Delta_0 \cup \Delta) &= (\Psi_1 * \Delta_0) * (\Delta_0 \cup \Delta) \\ &= (\Psi_2 * \Delta_0) * (\Delta_0 \cup \Delta) \\ &= \Psi_2 * (\Delta_0 \cup \Delta). \end{aligned}$$

Focusing and conditioning

Focusing means applying generic knowledge to a reference class appropriate to describe the context of interest [Dubois & Prade, 1996].

Focusing and conditioning

Focusing means applying generic knowledge to a reference class appropriate to describe the context of interest [Dubois & Prade, 1996].
$\rightarrow$ shift in context (to that reference class)
$\rightarrow$ update by a fact with probability 1

Focusing and conditioning

Focusing means applying generic knowledge to a reference class appropriate to describe the context of interest [Dubois & Prade, 1996].
$\rightarrow$ shift in context (to that reference class)
$\rightarrow$ update by a fact with probability 1
$\rightarrow$ via ME: conditioning the prior epistemic state.

Focusing and conditioning

Focusing means applying generic knowledge to a reference class appropriate to describe the context of interest [Dubois & Prade, 1996].
$\rightarrow$ shift in context (to that reference class)
$\rightarrow$ update by a fact with probability 1
$\rightarrow$ via ME: conditioning the prior epistemic state.

Indeed, conditioning is usually considered to be the proper operation for focusing, but has also been used for (general) revision [Gärdenfors, 1988].

Focusing and conditioning

Focusing means applying generic knowledge to a reference class appropriate to describe the context of interest [Dubois & Prade, 1996].
$\rightarrow$ shift in context (to that reference class)
$\rightarrow$ update by a fact with probability 1
$\rightarrow$ via ME: conditioning the prior epistemic state.

Indeed, conditioning is usually considered to be the proper operation for focusing, but has also been used for (general) revision [Gärdenfors, 1988].

However, focusing and revision differ conceptually: revision is not only applying knowledge, but means incorporating a new belief/constraint so as to change knowledge.

Focusing and conditioning

Focusing means applying generic knowledge to a reference class appropriate to describe the context of interest [Dubois & Prade, 1996].
$\rightarrow$ shift in context (to that reference class)
$\rightarrow$ update by a fact with probability 1
$\rightarrow$ via ME: conditioning the prior epistemic state.

Indeed, conditioning is usually considered to be the proper operation for focusing, but has also been used for (general) revision [Gärdenfors, 1988].

However, focusing and revision differ conceptually: revision is not only applying knowledge, but means incorporating a new belief/constraint so as to change knowledge.

Due to this conceptual mismatch, paradoxes have been observed (we'll come back to this later on).

The difference between focusing and revision

In our framework, it is easily possible to treat revision differently from focusing.

The following proposition reveals the difference between (conservative) revision by a certain information A, and focusing to A by conditioning:

Proposition

Let P be a distribution, $\mathcal{R} \subseteq (\mathcal{L} \mid \mathcal{L})^{prob}$ a (P-consistent[a]) set of probabilistic conditionals, and let $A[1]$ be a certain probabilistic fact.

(i) Focusing on A, i.e., updating P with $A[1]$ is done by ME-revision and yields $P *_{ME} \{A[1]\} = P(\cdot|A)$; in particular, $(P *_{ME} \mathcal{R}) *_{ME} A[1] = (P *_{ME} \mathcal{R})(\cdot|A)$.

(ii) Conservatively revising $P *_{ME} \mathcal{R}$ with $A[1]$ yields $P *_{ME} (\mathcal{R} \cup \{\mathrm{A}[1]\}) = P(\cdot|\mathrm{A}) *_{ME} \mathcal{R}$.

[a] $\mathcal{R}$ is P-consistent if there is a distribution Q with $Q \models \mathcal{R}$ and $Q(\omega) = 0$ whenever $P(\omega) = 0$.

Example psychologist (cont'd)

We illustrate the difference between focusing and (conservative) revision in the psychologist example.

Remember: The psychologist is now working in a clinic in which exclusively people being addicted to alcohol and/or drugs are treated, and the rate of people being addicted to alcohol, but also addicted to drugs amounts to 40 %.

His current epistemic state is given by $P^* = P *_{ME} \mathcal{R}$ with $\mathcal{R} = \{a \vee d[1], (d|a)[0.4]\}$:

ω	$P^*(\omega)$	ω	$P^*(\omega)$	ω	$P^*(\omega)$	ω	$P^*(\omega)$
ady	0.099	$\overline{a}dy$	0.425	$ad\overline{y}$	0.105	$\overline{a}d\overline{y}$	0.066
$a\overline{d}y$	0.089	$\overline{a}\overline{d}y$	0.0	$a\overline{d}\overline{y}$	0.216	$\overline{a}\overline{d}\overline{y}$	0.0

Example Psychologist (Cont'd)

After a few days, the psychologist notices that there are only young people in the new clinic; he has to revise his knowledge conservatively and computes $P *_{ME} (\mathcal{R} \cup y[1]) =: P_1^*$:

ω	$P_1^*(\omega)$	ω	$P_1^*(\omega)$
ady	0.120	$\overline{a}dy$	0.700
$ad\overline{y}$	0.0	$\overline{a}d\overline{y}$	0.0
$a\overline{d}y$	0.180	$\overline{a}\overline{d}y$	0.0
$a\overline{d}\overline{y}$	0.0	$\overline{a}\overline{d}\overline{y}$	0.0

We still have $P_1^* \models (d|a)[0.1]$ and $P_1^* \models (d|\overline{a})[1]$, since both conditionals are part of $\mathcal{R} \cup \{y[1]\}$, but now we find $P_1^*(a|y) = 0.3$ and $P_1^*(d|y) = 0.82$ – these probabilities have slightly changed.

Example Psychologist (Cont'd)

Note that this probability distribution is different from that which can be obtained via conditioning by y from $P^* = P *_{ME} \mathcal{R}$:

ω	$P^*(\cdot\|y)(\omega)$	ω	$P^*(\cdot\|y)(\omega)$
ady	0.162	$\overline{a}dy$	0.693
$ad\overline{y}$	0.0	$\overline{a}d\overline{y}$	0.0
$a\overline{d}y$	0.145	$\overline{a}\overline{d}y$	0.0
$a\overline{d}\overline{y}$	0.0	$\overline{a}\overline{d}\overline{y}$	0.0

These (conditional) probabilities are suitable for the case of applying the knowledge of the psychologist to a young person (focusing).

Example: Peter, Paul, and Mary

The following example (in various forms) has been used to blame MaxEnt for performing erroneous reasoning:

Example (Dubois, Prade, Smets)

Peter, Paul, and Mary are killers one of whom has been hired by Big Boss to commit a murder. Police Inspector Smith knows that Big Boss has first tossed a coin to decide whether it should be a male (Peter or Paul), or a female (Mary), but he does not know about the outcome of the tossing.

Example: Peter, Paul, and Mary

The following example (in various forms) has been used to blame MaxEnt for performing erroneous reasoning:

Example (Dubois, Prade, Smets)

Peter, Paul, and Mary are killers one of whom has been hired by Big Boss to commit a murder. Police Inspector Smith knows that Big Boss has first tossed a coin to decide whether it should be a male (Peter or Paul), or a female (Mary), but he does not know about the outcome of the tossing.

So, initially, the explicit beliefs of Smith are given by

$$\mathcal{R}_1 = \{(Peter \vee Paul)[0.5], Mary[0.5]\},$$

Example: Peter, Paul, and Mary

The following example (in various forms) has been used to blame MaxEnt for performing erroneous reasoning:

Example (Dubois, Prade, Smets)

Peter, Paul, and Mary are killers one of whom has been hired by Big Boss to commit a murder. Police Inspector Smith knows that Big Boss has first tossed a coin to decide whether it should be a male (Peter or Paul), or a female (Mary), but he does not know about the outcome of the tossing.

So, initially, the explicit beliefs of Smith are given by

$$\mathcal{R}_1 = \{(Peter \vee Paul)[0.5], Mary[0.5]\},$$

and his initial epistemic state can be calculated via the MaxEnt:

$$P_1 = ME(\mathcal{R}_1).$$

It is straightforward to see that
$P_1(Mary) = 0.5, P_1(Paul) = P_1(Peter) = 0.25.$

Example: Peter, Paul, and Mary (cont'd)

Example

Now Smith comes to know that Peter has been arrested right before the murder, so he could not have committed the crime.

Example: Peter, Paul, and Mary (cont'd)

Example

Now Smith comes to know that Peter has been arrested right before the murder, so he could not have committed the crime.

This piece of information can be encoded by

$$\mathcal{R}_2 = \{\neg Peter[1]\}.$$

Example: Peter, Paul, and Mary (cont'd)

Example

Now Smith comes to know that Peter has been arrested right before the murder, so he could not have committed the crime.

This piece of information can be encoded by

$$\mathcal{R}_2 = \{\neg Peter[1]\}.$$

It has been argued that $\mathcal{R}_2$ should be incorporated by updating (= conditioning). Doing so, the new epistemic state would be

$$P_2 = P_1(\cdot|\neg Peter),$$

and hence the new beliefs concerning Paul and Mary would be

$$P_2(Mary) = \frac{2}{3} \quad \text{and} \quad P_2(Paul) = \frac{1}{3}.$$

This seems to be unintuitive, as it gives undue precedence to Mary.

Example: Peter, Paul, and Mary (cont'd)

However, this flaw is neither an argument against MaxEnt, nor against probability theory in general, but caused by the confusion between focusing and revision.

Example: Peter, Paul, and Mary (cont'd)

However, this flaw is neither an argument against MaxEnt, nor against probability theory in general, but caused by the confusion between focusing and revision.

The correct change operation here is conservative revision, as $\mathcal{R}_2$ refers to the same context ($=$ murder case) as $\mathcal{R}_1$.

By conservative revision, we obtain

$$P_3 = ME(\mathcal{R}_1 \cup \mathcal{R}_2),$$

Example: Peter, Paul, and Mary (cont'd)

However, this flaw is neither an argument against MaxEnt, nor against probability theory in general, but caused by the confusion between focusing and revision.

The correct change operation here is conservative revision, as $\mathcal{R}_2$ refers to the same context (= murder case) as $\mathcal{R}_1$.

By conservative revision, we obtain

$$P_3 = ME(\mathcal{R}_1 \cup \mathcal{R}_2),$$

where (intuitively correct)

$$P_3(Mary) = P_3(Paul) = 0.5$$

holds.

This resolves the paradox.

Overview of this talk – Part II

- Induction and revision in probabilistics:
  - The entropy principles MaxEnt and MinCEnt
- Different revision scenarios: conservative revision vs. focusing
- Induction and revision with ranking functions:
  - C-representations and c-revisions
- Conclusion and outlook

Part I: Ranking functions and conditionals

A particular useful implementation of a plausibility relation:

Ordinal conditional functions (OCF, ranking functions[1]) [Spohn 1988]
$\kappa : \Omega \to \mathbb{N}(\cup\{\infty\})$ (Ω set of possible worlds, $\kappa^{-1}(0) \neq \emptyset$)

[1]Rankings can be understood as qualitative abstractions of probabilities

Part I: Ranking functions and conditionals

A particular useful implementation of a plausibility relation:

Ordinal conditional functions (OCF, ranking functions[1]) [Spohn 1988]
$\kappa : \Omega \to \mathbb{N}(\cup\{\infty\})$ (Ω set of possible worlds, $\kappa^{-1}(0) \neq \emptyset$)

$\kappa(\omega_1) < \kappa(\omega_2)$	ω_1 is more plausible than ω_2
$\kappa(\omega) = 0$	ω is maximally plausible

[1]Rankings can be understood as qualitative abstractions of probabilities

Part I: Ranking functions and conditionals

A particular useful implementation of a plausibility relation:

Ordinal conditional functions (OCF, ranking functions[1]) [Spohn 1988]

$\kappa : \Omega \to \mathbb{N}(\cup\{\infty\})$ (Ω set of possible worlds, $\kappa^{-1}(0) \neq \emptyset$)

$\kappa(\omega_1) < \kappa(\omega_2)$	ω_1 is more plausible than ω_2
$\kappa(\omega) = 0$	ω is maximally plausible
$\kappa(A)$	$:= \min\{\kappa(\omega) \mid \omega \models A\}$

[1] Rankings can be understood as qualitative abstractions of probabilities

Part I: Ranking functions and conditionals

A particular useful implementation of a plausibility relation:

Ordinal conditional functions (OCF, ranking functions[1]) [Spohn 1988]
$\kappa : \Omega \to \mathbb{N}(\cup\{\infty\})$ (Ω set of possible worlds, $\kappa^{-1}(0) \neq \emptyset$)

$\kappa(\omega_1) < \kappa(\omega_2)$	ω_1 is more plausible than ω_2
$\kappa(\omega) = 0$	ω is maximally plausible
$\kappa(A)$	$:= \min\{\kappa(\omega) \mid \omega \models A\}$
$\mathit{Bel}(\kappa)$	$:= \{A \mid \kappa(\neg A) > 0\}$

[1] Rankings can be understood as qualitative abstractions of probabilities

Part I: Ranking functions and conditionals

A particular useful implementation of a plausibility relation:

Ordinal conditional functions (OCF, ranking functions[1]) [Spohn 1988]

$\kappa : \Omega \rightarrow \mathbb{N}(\cup\{\infty\})$ (Ω set of possible worlds, $\kappa^{-1}(0) \neq \emptyset$)

$\kappa(\omega_1) < \kappa(\omega_2)$	ω_1 is more plausible than ω_2
$\kappa(\omega) = 0$	ω is maximally plausible
$\kappa(A)$	$:= \min\{\kappa(\omega) \mid \omega \models A\}$
$\mathit{Bel}(\kappa)$	$:= \{A \mid \kappa(\neg A) > 0\}$

Validating conditionals

$$\kappa \models (B|A) \text{ iff } \kappa(AB) < \kappa(A\overline{B}) \text{ iff } A \mid\!\sim_\kappa B$$

- κ accepts a conditional $(B|A)$
  - iff its verification AB is more plausible than its falsification $A\overline{B}$
  - iff from A, defeasibly infer B (based on κ).

[1] Rankings can be understood as qualitative abstractions of probabilities

C-revisions

Transferring the basic ideas underlying the ME-principles to the framework of ranking functions brings us to c-revisions and c-representations:

Let κ be a ranking function, $\Delta = \{(B_1|A_1), \ldots, (B_n|A_n)\}$ be a set of conditionals.

C-revisions

Transferring the basic ideas underlying the ME-principles to the framework of ranking functions brings us to c-revisions and c-representations:

Let κ be a ranking function, $\Delta = \{(B_1|A_1), \ldots, (B_n|A_n)\}$ be a set of conditionals.

C-revision $\kappa *_c \Delta$

A c-revision of κ by Δ is defined via

$$\kappa *_c \Delta(\omega) = \kappa_0 + \kappa(\omega) + \sum_{\substack{1 \leqslant i \leqslant n \\ \omega \models A_i \overline{B_i}}} \kappa_i^-$$

such that

$$\kappa_i^- > \min_{\omega \models A_i B_i} (\kappa(\omega) + \sum_{\substack{j \neq i \\ \omega \models A_j \overline{B}_j}} \kappa_j^-) - \min_{\omega \models A_i \overline{B}_i} (\kappa(\omega) + \sum_{\substack{j \neq i \\ \omega \models A_j \overline{B}_j}} \kappa_j^-),$$

where the second condition ensures that $\kappa *_c \Delta \models \Delta$.

C-representations

C-representations arise from applying c-revisions to the uniform ranking function κ_u[2]:

C-representations for inductive reasoning

A c-representation of Δ yields an inductive model of Δ, $\mathit{ind}(\Delta) = \kappa_\Delta$ which has the form

$$\kappa_\Delta(\omega) = \sum_{\substack{1 \leqslant i \leqslant n \\ \omega \models A_i \overline{B_i}}} \kappa_i^-$$

such that

$$\kappa_i^- > \min_{\omega \models A_i B_i} \Big(\sum_{\substack{j \neq i \\ \omega \models A_j \overline{B}_j}} \kappa_j^- \Big) - \min_{\omega \models A_i \overline{B}_i} \Big(\sum_{\substack{j \neq i \\ \omega \models A_j \overline{B}_j}} \kappa_j^- \Big).$$

[2] $\kappa_u(\omega) = 0$ for all ω

Different scenarios of induction and revision

C-representations and c-revisions behave very similarly to MaxEnt and MinCEnt, in particular, they allow for distinguishing among different reasoning and revision scenarios such as focusing vs. revision.

Different scenarios of induction and revision

C-representations and c-revisions behave very similarly to MaxEnt and MinCEnt, in particular, they allow for distinguishing among different reasoning and revision scenarios such as focusing vs. revision.

For this, we need to extend c-revisions by also dealing with adopting certain facts A^∞ by assigning ∞ to all falsifying worlds.

Different scenarios of induction and revision

C-representations and c-revisions behave very similarly to MaxEnt and MinCEnt, in particular, they allow for distinguishing among different reasoning and revision scenarios such as focusing vs. revision.

For this, we need to extend c-revisions by also dealing with adopting certain facts A^∞ by assigning ∞ to all falsifying worlds.

Exploring the schema of c-revisions for this case yields

$$\kappa *_c A^\infty(\omega) = \begin{cases} \kappa|A(\omega) & \text{if } \omega \models A \\ \infty & \text{if } \omega \not\models A. \end{cases}$$

Note the difference to revising by a plausible fact $(A|\top)$ which only makes falsifying worlds less plausible.

Revision vs. focusing

Proposition

Let κ be a ranking function, $\Delta \subseteq (\mathcal{L} \mid \mathcal{L})$ a (κ-consistent[a]) set of conditionals, and suppose A to be a certain fact.

(i) Focusing κ on A, i.e., updating κ with the certain fact A via c-revision is done by conditioning and yields $\kappa *_c A^\infty(\omega) = \kappa|A(\omega)$ for models ω of A; in particular, $(\kappa *_c \Delta) *_c A^\infty = (\kappa *_c \Delta)|A$ on the models of A.

(ii) Conservatively revising $\kappa *_c \Delta$ with the certain fact A yields $\kappa *_c (\Delta \cup \{\mathrm{A}^\infty\}) = (\kappa *_c \mathrm{A}^\infty) *_c \Delta$ (which coincides with $(\kappa|A) *_c \Delta$ on the models of A) if the same parameters κ_i^- are chosen for both c-revisions.

[a] Δ is κ-consistent if there is a ranking function κ' with $\kappa' \models \Delta$ and $\kappa'(\omega) = \infty$ whenever $\kappa(\omega) = \infty$.

The psychologist's example

We generate an inductive model $\kappa = \kappa_{psych}$ containing the conditionals

$$\Delta_{psych} = \{(\overline{d}|a), (\overline{a}|d), (\overline{a}|y), (d|y), (\overline{d}|\overline{y}), (a|\overline{y})\}$$

(and more) by choosing a c-representation with pareto-minimal parameters κ_i^-.

ω	$\kappa(\omega)$	$\kappa_1^*(\omega)$	$\kappa *_c y^\infty(\omega)$	$\kappa_2^*(\omega)$	$\kappa_3^*(\omega)$
ady	4	4	3	2	2
$ad\overline{y}$	4	4	∞	∞	∞
$a\overline{d}y$	3	3	2	1	1
$a\overline{d}\overline{y}$	0	0	∞	∞	∞
$\overline{a}dy$	1	6	0	3	4
$\overline{a}d\overline{y}$	4	9	∞	∞	∞
$\overline{a}\overline{d}y$	2	2	1	0	0
$\overline{a}\overline{d}\overline{y}$	3	3	∞	∞	∞

We adapt the story to the framework of ranking functions.

The psychologist's example (cont'd)

Now the psychologist is going to change his job: He will be working in a clinic where addictions to both alcohol and drugs are not uncommon, more precisely, people being addicted to drugs tend to also being addicted to alcohol.

The psychologist's example (cont'd)

Now the psychologist is going to change his job: He will be working in a clinic where addictions to both alcohol and drugs are not uncommon, more precisely, people being addicted to drugs tend to also being addicted to alcohol.

So, when starting to work in the new environment, the psychologist c-revises his initial epistemic state κ by $\Delta_1 = \{(a|d)\}$, yielding $\kappa_1^* = \kappa *_c \Delta_1$ (with minimal parameter).

The psychologist's example (cont'd)

$\kappa_1^* = \kappa *_c \Delta_1$ with $\Delta_1 = \{(a|d)\}$:

ω	$\kappa(\omega)$	$\kappa_1^*(\omega)$	$\kappa *_c y^\infty(\omega)$	$\kappa_2^*(\omega)$	$\kappa_3^*(\omega)$
ady	4	4	3	2	2
$ad\overline{y}$	4	4	∞	∞	∞
$a\overline{d}y$	3	3	2	1	1
$a\overline{d}\overline{y}$	0	0	∞	∞	∞
$\overline{a}dy$	1	6	0	3	4
$\overline{a}d\overline{y}$	4	9	∞	∞	∞
$\overline{a}\overline{d}y$	2	2	1	0	0
$\overline{a}\overline{d}\overline{y}$	3	3	∞	∞	∞

The psychologist's example (cont'd)

After having spent a couple of days in the new clinic, the psychologist realized that this clinic is for young people only, i.e., y^∞ holds.

The psychologist's example (cont'd)

After having spent a couple of days in the new clinic, the psychologist realized that this clinic is for young people only, i.e., y^∞ holds.

He conservatively revises κ_1^* by y^∞, yielding $\kappa_2^* = \kappa *_c \Delta_2$ with $\Delta_2 = \{(a|d), y^\infty\}$.

The psychologist's example (cont'd)

After having spent a couple of days in the new clinic, the psychologist realized that this clinic is for young people only, i.e., y^∞ holds.

He conservatively revises κ_1^* by y^∞, yielding $\kappa_2^* = \kappa *_c \Delta_2$ with $\Delta_2 = \{(a|d), y^\infty\}$.

Hence, according to the proposition above, he obtains $\kappa_2^* = (\kappa *_c y^\infty) *_c \Delta_1$.

The psychologist's example (cont'd)

After having spent a couple of days in the new clinic, the psychologist realized that this clinic is for young people only, i.e., y^∞ holds.

He conservatively revises κ_1^* by y^∞, yielding $\kappa_2^* = \kappa *_c \Delta_2$ with $\Delta_2 = \{(a|d), y^\infty\}$.

Hence, according to the proposition above, he obtains $\kappa_2^* = (\kappa *_c y^\infty) *_c \Delta_1$.

Note that κ_2^* is different from the ranking function that the psychologist would have obtained by focusing his beliefs represented by κ_1^* on a young person.

The psychologist's example (cont'd)

After having spent a couple of days in the new clinic, the psychologist realized that this clinic is for young people only, i.e., y^∞ holds.

He conservatively revises κ_1^* by y^∞, yielding $\kappa_2^* = \kappa *_c \Delta_2$ with $\Delta_2 = \{(a|d), y^\infty\}$.

Hence, according to the proposition above, he obtains $\kappa_2^* = (\kappa *_c y^\infty) *_c \Delta_1$.

Note that κ_2^* is different from the ranking function that the psychologist would have obtained by focusing his beliefs represented by κ_1^* on a young person.

In that case, he would have updated κ_1^* by y^∞, yielding $\kappa_3^* = \kappa_1^* *_c y^\infty$ (which coincides with $\kappa_1^*|y$ on the models of y).

The psychologist's example (cont'd)

After having spent a couple of days in the new clinic, the psychologist realized that this clinic is for young people only, i.e., y^∞ holds.

He conservatively revises κ_1^* by y^∞, yielding $\kappa_2^* = \kappa *_c \Delta_2$ with $\Delta_2 = \{(a|d), y^\infty\}$.

Hence, according to the proposition above, he obtains $\kappa_2^* = (\kappa *_c y^\infty) *_c \Delta_1$.

Note that κ_2^* is different from the ranking function that the psychologist would have obtained by focusing his beliefs represented by κ_1^* on a young person.

In that case, he would have updated κ_1^* by y^∞, yielding $\kappa_3^* = \kappa_1^* *_c y^\infty$ (which coincides with $\kappa_1^*|y$ on the models of y).

Clearly, κ_2^* and κ_3^* are different (though the differences are only small).

The psychologist's example (cont'd)

$\kappa_1^* = \kappa *_c \Delta_1$ with $\Delta_1 = \{(a|d)\}$

$\kappa_2^* = \kappa *_c \Delta_2 = (\kappa *_c y^\infty) *_c \Delta_1$ with $\Delta_2 = \{(a|d), y^\infty\}$

$\kappa_3^* = \kappa_1^* *_c y^\infty$

ω	$\kappa(\omega)$	$\kappa_1^*(\omega)$	$\kappa *_c y^\infty(\omega)$	$\kappa_2^*(\omega)$	$\kappa_3^*(\omega)$
ady	4	4	3	2	2
$ad\overline{y}$	4	4	∞	∞	∞
$a\overline{d}y$	3	3	2	1	1
$a\overline{d}\overline{y}$	0	0	∞	∞	∞
$\overline{a}dy$	1	6	0	3	4
$\overline{a}d\overline{y}$	4	9	∞	∞	∞
$\overline{a}\overline{d}y$	2	2	1	0	0
$\overline{a}\overline{d}\overline{y}$	3	3	∞	∞	∞

Overview of this talk – Part II

- Induction and revision in probabilistics:
  - The entropy principles MaxEnt and MinCEnt
- Different revision scenarios: conservative revision vs. focusing
- Induction and revision with ranking functions:
  - C-representations and c-revisions
- Conclusion and outlook

Conclusion

- We described inductive reasoning from conditional belief bases in a rich epistemic framework that takes epistemic states and conditionals as basic encodings of information.

Conclusion

- We described inductive reasoning from conditional belief bases in a rich epistemic framework that takes epistemic states and conditionals as basic encodings of information.
- Allowing inductive reasoning from background beliefs lead us naturally to consider also belief revision. Our main claim here is that inductive reasoning can be considered as a special case of epistemic belief revision.

Conclusion

- We described inductive reasoning from conditional belief bases in a rich epistemic framework that takes epistemic states and conditionals as basic encodings of information.
- Allowing inductive reasoning from background beliefs lead us naturally to consider also belief revision. Our main claim here is that inductive reasoning can be considered as a special case of epistemic belief revision.
- In this way, a coherent and homogeneous approach to inductive reasoning is possible that allows us to realize different forms of inductive reasoning via revision, updating, and focusing.

Conclusion

- We described inductive reasoning from conditional belief bases in a rich epistemic framework that takes epistemic states and conditionals as basic encodings of information.
- Allowing inductive reasoning from background beliefs lead us naturally to consider also belief revision. Our main claim here is that inductive reasoning can be considered as a special case of epistemic belief revision.
- In this way, a coherent and homogeneous approach to inductive reasoning is possible that allows us to realize different forms of inductive reasoning via revision, updating, and focusing.
- We presented a proof of concept via the probabilistic ME-principles and c-representations/c-revisions. Both framework rely crucially on an abstract/algebraic conditional-logical principle of conditional preservation [GKI, AMAI 2004] which subsumes the D&P principle [GKI, KR 2018].

A Novel Vision of Revision

Both AGM belief revision and DP[3] iterated revision are not enough for our framework because

- we need revision by sets of conditionals;
- we need to distinguish between background beliefs and contextual information;
- we need to distinguish between explicit and implicit beliefs.

[3] [Darwiche & Pearl, AIJ 1997]

A Novel Vision of Revision

Both AGM belief revision and DP[3] iterated revision are not enough for our framework because

- we need revision by sets of conditionals;
- we need to distinguish between background beliefs and contextual information;
- we need to distinguish between explicit and implicit beliefs.

$\rightarrow$ We need to talk about something like inductive belief revision here.

So, methods of induction may also lead to novel perspectives of revision.

[3] [Darwiche & Pearl, AIJ 1997]