Promotionskolloquium

Algebraic Uncertainty Theory

A Unifying Perspective on Reasoning under Uncertainty

Jörg Zimmermann

Institute of Computer Science
University of Bonn, Germany

Motivation and Context: Learning Systems

The search for foundations of effective learning systems leads to three main questions:

1. **How should a learning system represent and process uncertain information, or, what is the proper inductive logic?**
2. **What set of possible models should the system consider?**
3. **How to relate the explanatory power of a model to its complexity?**

How to represent and process Uncertainty?

- There are a lot of uncertainty calculi around today:

  Dempster-Shafer theory, possibility theory, ranking theory, revision theory, ...

- All of them make direct assumptions on the structure of uncertainty values (e.g. one real number).

- But why should uncertainty not be measured by two (like in DS-theory) or more real numbers, or by a complex number, or matrices, or ...

$\Rightarrow$ Derive structure of uncertainty from axioms, instead of defining it.

Why not just use Probability Theory: the Ellsberg Paradox

- **An urn contains 30 red balls and 60 other balls that are either black or yellow.**
- **First alternative:**

  Gamble A: You receive \$100 if you draw a red ball ($p = \frac{1}{3}$).

  Gamble B: You receive \$100 if you draw a black ball ($p \in [0, \frac{2}{3}]$).
- **Second alternative:**

  Gamble C: You receive \$100 if you draw a red or yellow ball ($p \in [\frac{1}{3}, 1]$).

  Gamble D: You receive \$100 if you draw a black or yellow ball ($p = \frac{2}{3}$).

Why not just use Probability Theory: the Ellsberg Paradox

Most people surveyed have the following preferences:

Gamble A $>$ Gamble B and Gamble D $>$ Gamble C.

But decision theory based on probability theory implies:

If Gamble A $>$ Gamble B, then Gamble C $>$ Gamble D.

- **This contradiction indicates that some assumptions of decision theory based on probability are violated.**
- **Point-probabilities can't distinguish between randomness and ignorance.**
- **The Ellsberg Paradox can be explained by "ambiguity aversion", which can't be modeled by probability theory.**

Terminology and Notation

- The domain of confidence values $\mathcal{C}$ is partially ordered and has a greatest ($\top$, "certain") and a least ($\perp$, "impossible") element.

- A conditional confidence measure for a Boolean Algebra $\mathbf{U}$ and a domain of confidence values $\mathcal{C}$ is a mapping:

$$\Gamma : \mathbf{U} \times \mathbf{U} \setminus \{\perp\} \to \mathcal{C}.$$

  Instead of $\Gamma(A, B)$ we will write $\Gamma(A|B)$ and say "the confidence value of A given B (wrt. Γ)".

- A confidence space is a triple $(\mathbf{U}, \Gamma, \mathcal{C})$.

- A set of confidence spaces sharing the same domain of confidence values we will call a confidence universe.

The Axiom System of R. T. Cox (1946)

1. $\mathcal{C} \subseteq \mathbb{R}$ **(i.e., confidence values are real numbers)**
2. **There is a function S:** $\Gamma(\neg A) = S(\Gamma(A))$.
3. **There is a function F:** $\Gamma(AB) = F(\Gamma(A|B), \Gamma(B))$.
4. **S, F are twice differentiable.**

Cox showed that confidence measures satisfying these axioms are effectively probability measures (but his axioms are incomplete!).

The Axiom System of S. Arnborg and G. Sjödin (2001)

Arnborg and Sjödin extend the axiom system of Cox with the Refinability Axiom (in total they need 16 axioms, but drop the real value assumption):

For every confidence space $(\mathbf{U}, \Gamma, \mathcal{C})$, it must be possible to introduce a new subcase B of a proposition A with confidence value v given to $\Gamma(B|A)$,

and two other conditions regarding the extendability of a proposition algebra.

New Axiom System NC$_{12}$: 3 Connective Axioms

(Not) **If** $\Gamma_1(A_1) = \Gamma_2(A_2)$**, then** $\Gamma_1(\neg A_1) = \Gamma_2(\neg A_2)$**.**

(And$_1$) **If** $\Gamma_1(A_1|B_1) = \Gamma_2(A_2|B_2)$ **and** $\Gamma_1(B_1) = \Gamma_2(B_2)$**,**
then $\Gamma_1(A_1 B_1) = \Gamma_2(A_2 B_2)$**.**

and axiom And$_2$.

New Axiom System $\mathbf{NC}_{12}$: 3 Infrastructure Axioms

(Extensibility) $\forall\ (\mathbf{U_1}, \Gamma_1, \mathcal{C})$ **and** $(\mathbf{U_2}, \Gamma_2, \mathcal{C})$ **there is a confidence space** $(\mathbf{U_3}, \Gamma_3, \mathcal{C})$**, so that** $\mathbf{U_3} \cong \mathbf{U_1} \otimes \mathbf{U_2}$**, and for all** $A_1, B_1 \in \mathbf{U}_1$**,** $A_2, B_2 \in \mathbf{U}_2$**:**

$$\Gamma_3(A_1 \otimes \top_2 \,|\, B_1 \otimes B_2) = \Gamma_1(A_1|B_1),$$

$$\Gamma_3(\top_1 \otimes A_2 \,|\, B_1 \otimes B_2) = \Gamma_2(A_2|B_2).$$

and two order-theoretic axioms, Order$_1$ and Order$_2$.

Ring Theorem

The domain of confidence values $\mathcal{C}$ of a confidence universe satisfying the axiom system NC$_{12}$ can be embedded in a partially ordered commutative ring $(\hat{\mathcal{C}}, 0, 1, \oplus, \odot, \leq)$.

Let $\hat{\cdot} : \mathcal{C} \to \hat{\mathcal{C}}$ be the embedding map, then the following holds:

$$\hat{\perp\!\!\!\perp} = 0, \quad \hat{\top\!\!\!\top} = 1,$$

$$\forall v, w \in \mathcal{C} : \; v \leq w \;\Leftrightarrow\; \hat{v} \leq \hat{w}.$$

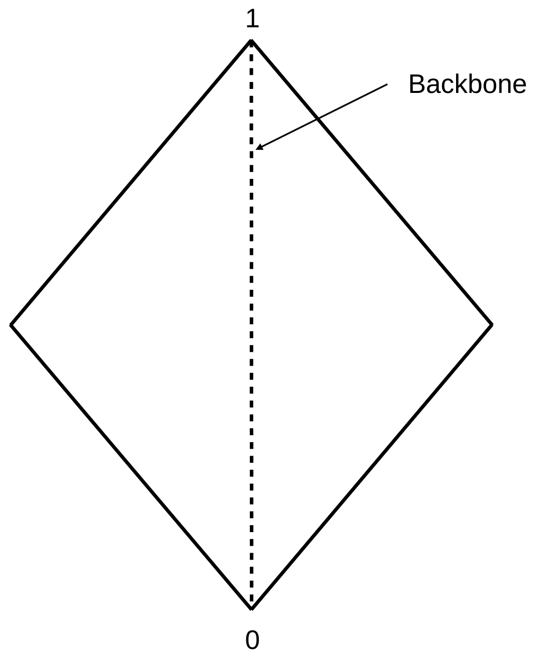
Ring Theorem

Furthermore, all confidence measures Γ of the confidence universe satisfy:

$$
\begin{aligned}
\hat{\Gamma}(\top) &= 1 \\
\hat{\Gamma}(A \vee B) &= \hat{\Gamma}(A) \oplus \hat{\Gamma}(B) \qquad\qquad \text{if } AB = \perp \\
\hat{\Gamma}(AB) &= \hat{\Gamma}(A|B) \odot \hat{\Gamma}(B)
\end{aligned}
$$

These properties are the analogs to the Kolmogorov axioms and conditionalization in probability theory, but now defined on a ring structure.

Decomposition of Uncertainty



The greatest totally ordered subfield of a confidence ring is called backbone (if it exists).

Backbone elements can be seen as numerical entities.

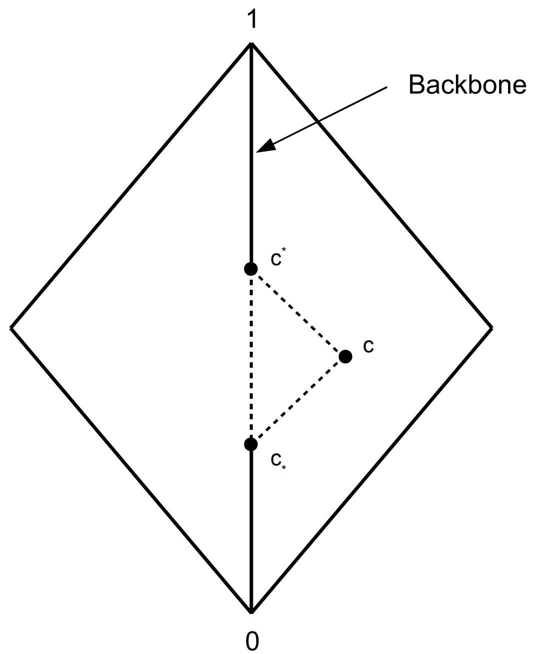
Decomposition Theorem

Given some weak additional assumption on confidence rings, every element of the $[0,1]$-interval can be decomposed into a numerical part and an interaction part:

$$\mathbf{c} = b + r \cdot \mathbf{a}$$

- $\mathbf{c}$ **is an arbitrary element of the** $[0,1]$**-interval.**
- b **and** r **are elements of the backbone.**
- $\mathbf{a}$ **is an "interaction element":** $0 \leq \mathbf{a} \leq 1$ **are the only bounds by backbone elements.**

Decomposition of Uncertainty



$$\mathbf{c}_* = b, \quad \mathbf{c}^* = b + r$$

Decomposition of Uncertainty

This decomposition result can be interpreted in the following way:

A general uncertainty value can be decomposed into a numerical interval $[b, b + r]$ and an interaction component $\mathbf{a}$.

If the interaction information can be neglected, this implies that uncertainty in general can be represented by numerical intervals.

Ellsberg Paradox revisited

Gamble A (red ball): $p = \frac{1}{3}$

Gamble B (black ball): $p = \frac{2}{3}\alpha = 0 + \frac{2}{3}\alpha,\ \ \alpha \in [0, 1]$

Decomposition: $b = 0, r = \frac{2}{3}, \mathbf{a} = \alpha$

Gamble C (red or yellow ball): $p = \frac{1}{3} + \frac{2}{3}\beta$

Decomposition: $b = \frac{1}{3}, r = \frac{2}{3}, \mathbf{a} = \beta$

Gamble D (black or yellow ball): $p = \frac{2}{3}\ \ (= \frac{2}{3}\alpha + \frac{2}{3}(1 - \alpha))$

Ellsberg Paradox revisited

Utility of Gamble A: $\frac{1}{3}U(\$100)$

Utility of Gamble B: $\frac{2}{3}\alpha \cdot U(\$100) \;\rightsquigarrow\; [0, \frac{2}{3}U(\$100)]$

Utility of Gamble C: $(\frac{1}{3} + \frac{2}{3}\beta) \cdot U(\$100) \;\rightsquigarrow\; [\frac{1}{3}U(\$100), U(\$100)]$

Utility of Gamble D: $\frac{2}{3}U(\$100)$

Most people seem to assume worst case when faced with lack of knowledge.

Relations to Dempster-Shafer Theory: Representation Theorem

For all Belief functions Bel (over finite proposition algebras) there is a confidence measure Γ with:

$$\Gamma_*(A) = Bel(A)$$

Γ_* returns the lower bound of the numerical interval defined by the decomposition of a confidence value.

So, roughly one can say that Dempster-Shafer theory is confidence theory without interaction information.

Unifying Power of Confidence Theory

Probability Theory:

+ Coherent Conditionalization

- Can't resolve Ellsberg Paradox

Dempster-Shafer Theory:

- No coherent Conditionalization

+ Resolution of Ellsberg Paradox

Confidence Theory:

+ Coherent Conditionalization

+ Resolution of Ellsberg Paradox

Effective Learning Systems

- **Learning in program space: Solomonoff induction**
- **In general not effective, but natural effective instances can be defined by combining search in program space and proof space (reduction of learnability to provability).**
- **Solomonoff induction doesn't tell the learning system whether the generation of the next bit has taken 1 second or 100 billion years. This is the cause of its incomputability.**
- **If the learning system is enhanced by an internal clock, all effectively generated bit sequences are effectively learnable.**

Publications

- **Jörg Zimmermann and Armin B. Cremers:**

  The Quest for Uncertainty

  C. Calude, G. Rozenberg, A. Salomaa (Eds.): Rainbow of Computer Science, pp. 270-283, Springer, 2011

- **Jörg Zimmermann and Armin B. Cremers:**

  Making Solomonoff Induction Effective or You Can Learn What You Can Bound

  S. B. Cooper, A. Dawar, B. Löwe (Eds.): How the World Computes, pp. 745-754, LNCS 7318, Proceedings of the CiE 2012, Turing Centenary Conference, Springer, 2012

Conclusions

1. **Algebraic Uncertainty Theory enables a unifying perspective on reasoning under uncertainty by deriving, and not defining, the structure of uncertainty values – it is not a YAUC (yet another uncertainty theory).**

2. **Algebraic Uncertainty Theory can solve longstanding problems like combining coherent conditionalization with a resolution of the Ellsberg Paradox.**

3. **If the learning system is enhanced by an internal clock:**

   Effective universal induction is possible!