Euclidean Properties of Bayesian Updating

Kyle Chauvin
Discussed by Andy Haupt

MURI Working Group

July 2020

Setting

- Consider Bayesian learning of one agent
- Beliefs with updating as an algebraic and geometric structure
- Consider "learning rules up to renaming", isomorphism classes

Log-Odds Ratio

- π prior for $x \in \mathcal{X}$. Define

$$o(\pi)(x|x_o) := \log \frac{\pi(x)}{\pi(x_0)} \qquad l(y)(x|x_0) := \log \frac{\pi(y|x)}{\pi(y|x_0)}.$$

- Then, Bayes' rule to update π with observation y to $y.\pi$ is

$$o(y.\pi)(x|x_0) = o(\pi)(x|x_0) + l(y)(x|x_0)$$

- Bayes' rule is "isomorphic to" addition of likelihood vector
- Encode log-odds with one number in $\mathbb{R}$: What learning rules admit such a representation?
- Consider geometry of function $\gamma(y, y') := \langle l(y), l(y') \rangle$

Some Algebra Notation

- $(\mathcal{A}, \cdot)$ is a semigroup iff $\cdot \colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ is associative
- $\bullet.\bullet \colon \mathcal{A} \times \mathcal{B} \to \mathcal{B}, (a, b) \mapsto a.b$ is a semigroup action iff $(a \cdot b).c = a.b.c$ for $a, b \in \mathcal{A}$, $c \in \mathcal{B}$.
- A semigroup action isomorphism $(\mathcal{A}, \mathcal{B}) \cong (\mathcal{A}', \mathcal{B}')$ is a pair of bijective functions $g \colon \mathcal{A} \to \mathcal{A}'$, $h \colon \mathcal{B} \to \mathcal{B}'$ such that $h(a.b) = g(a).h(b)$ for any $a \in \mathcal{A}$, $b \in \mathcal{B}$.

Learning Rules

Learning Rule

A Learning Rule is a semigroup action $\mathcal{A} \times \mathcal{B} \to \mathcal{B}$

- $a \in \mathcal{A}$ are called arguments (tacitly assume as countable)
- $b \in \mathcal{B}$ are called beliefs

Definition (Log-Bayesian Learning Rule)

$(\Omega, \mathcal{F}, \mathbb{P})$ discrete probability space, $Y_i \colon \Omega \to \mathcal{X}$, $i \in \mathbb{N}$, X, Y_i conditionally iid given X, $p(\bullet|\bullet)$ conditional pmf.
Define $y.\pi(x) := \frac{\pi(x)p(y|x)}{\sum_{x'} p(y|x')}$

Virtual Bayesian

A Learning Rule is Virtually Bayesian if it is isomorphic to the Bayesian Learning Rule.

Theorem

$(\mathcal{A}, \mathcal{B})$ is virtually Bayesian iff

Self-Recording *There is b_0 (ur-prior) such that $a \mapsto a.b_0$ is bijective,*

- *$\mathcal{A}$ is commutative, acyclic, and has rank at least 2*
- *$b \mapsto a.b$ is injective*

Lemma

Self-recording learning rules are isomorphic iff their argument semigroups are isomorphic.

Non-Examples

- Self-recording: Beliefs are too rich
- Commutativity: Rejecting information too far from one's prior
- Acyclicity: The summer is warm if it did not rain yesterday
- Rank ≥ 2: A memoryless counter
- Injectivity: I don't ever loose memory

De Groot

- De Groot is non-commutative, hence not virtually Bayesian

Proof Strategy

- It suffices to find a semigroup isomorphism for $\mathcal{A}$
- By commutativity and self-recording, there is an identity element in $\mathcal{A} \Rightarrow \mathcal{A}$ is a monoid
- A commutative monoid can be embedded into an Abelian Grothendieck group
- Any countable acyclic Abelian group can be embedded into a finite-dimensional $\mathbb{Q}$-vector space with basis $\{b_1, b_2, \ldots, b_n\}$.
- Let $X = \{x_1, x_2, \ldots, x_n\} \subseteq \mathbb{R}$ be $\mathbb{Q}$-independent, $\mathbb{R}_{<0} \cap X \neq \emptyset$. Consider the embedding $\sum_{i=1}^n q_i x_i \mapsto \sum_{i=1}^n q_i b_i$.
- Observe that this gives an embedding $\mathcal{A} \hookrightarrow \mathbb{R}$
- Construct binary Bayesian learner with matching log likelihood

Algebra of updating

- Agent has prior probability π on the alternative that a coin shows head w.p. $q \neq \frac{1}{2}$ (with alternative that it shows tail w.p. q).
  - It is isomorphic (a relabeling) to use natural language
    - $\pi = \frac{1}{2}$: The coin is neutral
    - $\pi = q$: Coin is heads-biased
    - $\pi = \frac{q^2}{q^2+(1-q)^2}$: Coin is very biased
    - $\pi = \frac{q^{n+1}}{q^{n+1}+(1-q)^{n+1}}$: Coin is $\underbrace{\text{very, \dots, very}}_{n}$ biased
  - Bernoulli Bayesian has $\mathcal{A} \cong \mathbb{N}$
  - Beta Bayesian has $\mathcal{A} \cong \mathbb{N} \times \mathbb{N}$

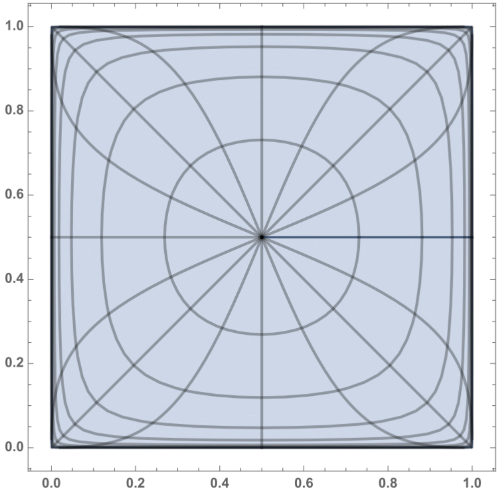
Theorem

A learning rule is isomorphic to the Bayesian lerning rule iff it is self-recording, has at least rank 2 and admits a bi-additive, symmetric, positive definite function $\gamma\colon \mathcal{A} \times \mathcal{A} \to \mathbb{R}$.

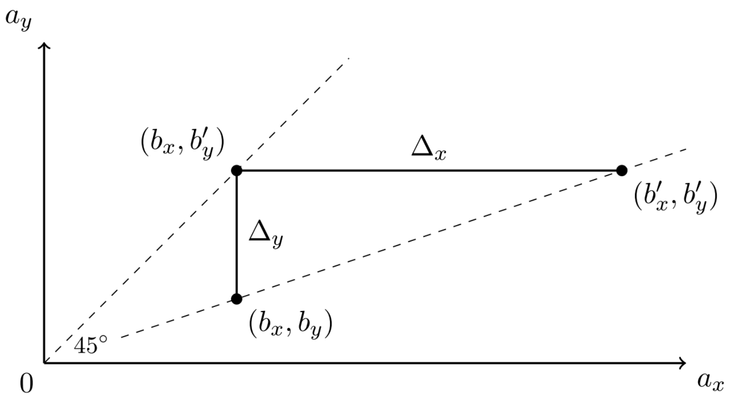
Theorem

- *For any bi-additive, symmetric, positive definite function $\gamma\colon \mathcal{A} \times \mathcal{A} \to \mathbb{R}$, there is $n \in \mathbb{N}$ and an essential embedding $\mathcal{A} to \mathbb{R}^n$ unique up to orthogonal transformations such that $\gamma(a, b) = \langle f(a), f(b) \rangle$.*
- *For any essential embedding $f : \mathcal{A} \to \mathbb{R}^n$ there is a unique bi-additive, symmetric, positive definite function function $\gamma\colon \mathcal{A} \times \mathcal{A} \to \mathbb{R}$.*

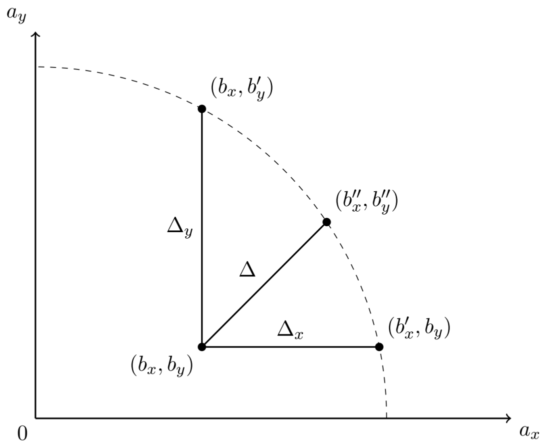
$\to$ Define angle, length, projection.


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Dynamic Belief Elicitation



Dynamic Belief Elicitation II



Discussion & Open Questions

Main Takeaways

- Many learning rules are isomorphic to Bayesian learning (Virtually Bayesian rules)
- Virtually Bayesian rules can be embedded into a finite-dimensional real vector space
- We can identify prior by dynamic experiments

Future Challenges

1. Classification of Bayesian isomorphism classes
2. Computing for
3. Modelling cognitive biases in geometric terms.