Towards a Strictly Frequentist Theory of Imprecise Probability

Christian Fröhlich[1], Rabanus Derr[1], Robert C. Williamson[1,2]

[1]University of Tübingen, [2]Tübingen AI Center.

The FMLS group

Foundations of Machine Learning Systems

Goal: to develop new and better foundations for machine learning (ML) systems

Research themes:

- Non-linear expectations and imprecise probability
- Interaction of probability theory and open problems in ML
- The style of reasoning of ML
- Information processing equalities

New models for data in ML

An overarching agenda

Data has many imperfections:

- Data corruptions
- Dataset shift: not a single, stable distribution

$\Longrightarrow$ need to move beyond the **i.i.d.** assumption.

Our interest: to account for failure of *statistical stability*.

Strict frequentism + divergence

- Aim: principled way to model **unstable** data.
- Strict frequentism: **sequence** as the primitive – cf. von Mises collectives.
- When the "probability" does not exist, natural replacement is an **upper prevision**.
  - Convergent case: probability is limiting relative frequency.
  - Now: cluster points of relative frequencies yield coherent upper prevision.
- Works for *all* sequences and *all* events/gambles.
- Conversely, to any coherent upper prevision can construct a corresponding sequence.
- Conditioning principle: recover the **generalized Bayes rule**.
- Independence is subtle, requires paying attention to set systems.

Technical Setup

Our setting

- Let Ω be an arbitrary set of outcomes. A *gamble* is a bounded function $X\colon \Omega \to \mathbb{R}$.
- We assume a loss-based orientation: $+$ is loss, $-$ is gain.
- The set L^∞ of bounded gambles forms a Banach space, with dual $(L^\infty)^*$ on which we install the weak* topology.
- The set of linear previsions is compact under the weak* topology:

$$\{E \in (L^\infty)^*\colon E(X) \geq 0 \text{ if } X \geq 0, E(\chi_\Omega) = 1\} \subset (L^\infty)^*.$$

- Let $\vec{\Omega}\colon \mathbb{N} \to \Omega$, an Ω-valued sequence of outcomes.
- Define sequence of linear previsions: $\vec{E}(n) := X \mapsto \frac{1}{n}\sum_{i=1}^n X\left(\vec{\Omega}(i)\right)$.

The forward direction

From sequence to upper prevision

Define a coherent upper (resp. lower) prevision as:

$$\overline{R}(X) := \sup\left\{E(X)\colon E \in \mathrm{CP}(\vec{E})\right\}, \quad \forall X \in L^\infty; \ \underline{R}(X) := -\overline{R}(-X).$$

Proposition:

$$\mathrm{CP}\left(n \mapsto \frac{1}{n}\sum_{i=1}^n X\left(\vec{\Omega}(i)\right)\right) = \left\{E(X)\colon E \in \mathrm{CP}(\vec{E})\right\}, \quad \forall X \in L^\infty.$$

Therefore:

$$\overline{R}(X) = \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X\left(\vec{\Omega}(i)\right).$$

- When the limiting relative frequency does not exist, use the $\limsup$ for decision making! (due to loss-based orientation)

The converse direction

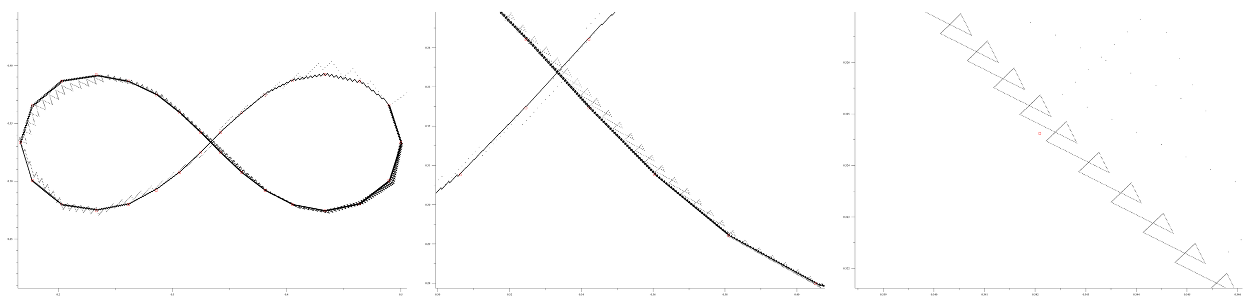
From upper prevision to sequence

Theorem: Let $|\Omega| < \infty$. Let $\overline{R}$ be a coherent upper prevision on L^∞. There exists a sequence $\vec{\Omega}$ such that we can write $\overline{R}$ as ($\forall X \in L^\infty$):

$$\overline{R}(X) = \sup\left\{E(X)\colon E \in \mathcal{E}_{\vec{\Omega}}\right\}, \ \mathcal{E}_{\vec{\Omega}} := \mathrm{CP}\left(\vec{E}_{\vec{\Omega}}\right),$$

where $\vec{E}_{\vec{\Omega}}(n) = X \mapsto \frac{1}{n}\sum_{i=1}^n X\left(\vec{\Omega}(i)\right)$, $\forall X \in L^\infty$.

- Strictly frequentist semantics for coherent upper previsions.
- Proof is constructive! Proof idea, visually:



- Walley & Fine [6] offer similar result for upper probabilities.

Conditional upper prevision

Inspired by von Mises

Sequence of conditional linear previsions (χ_B is the indicator function of B):

$$\vec{E}(\cdot|B)(n) := X \mapsto \sum_{i=1}^n (X \cdot \chi_B)\left(\vec{\Omega}(i)\right) \Big/ \sum_{i=1}^n \chi_B\left(\vec{\Omega}(i)\right).$$

Define the **conditional upper prevision**:

$$\overline{R}(X|B) := \sup\left\{E(X)\colon E \in \mathrm{CP}\left(\vec{E}(\cdot|B)\right)\right\}, \ \forall X \in L^\infty.$$

For $\underline{R}(\chi_B) > 0$, define the conditional set of desirable gambles as:

$$\mathcal{D}_{\vec{\Omega}|B} := \left\{X \in L^\infty\colon \overline{R}(X\chi_B) \leq 0\right\},$$

and a corresponding upper prevision, the generalized Bayes rule, as:

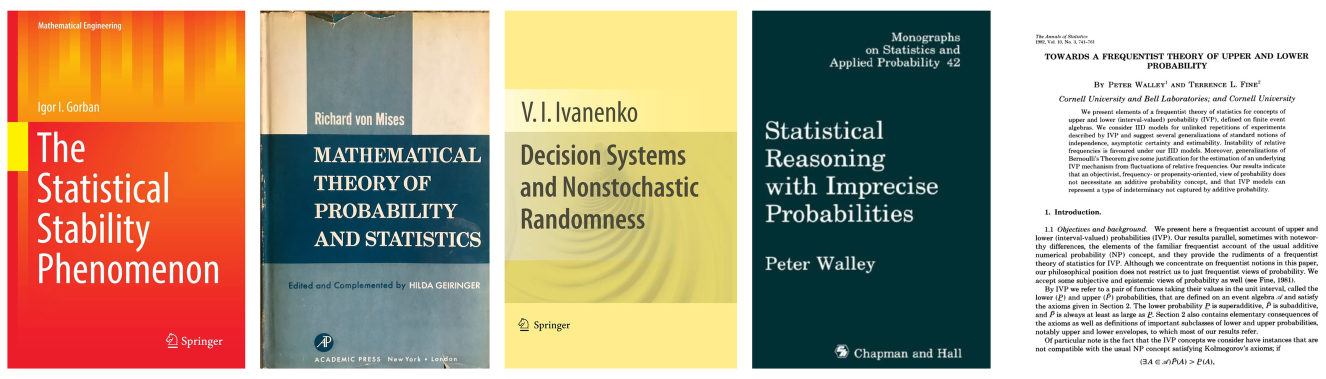
$$\mathrm{GBR}(X|B) := \inf\left\{\alpha \in \mathbb{R}\colon X - \alpha \in \mathcal{D}_{\vec{\Omega}|B}\right\}.$$

Proposition: If $\underline{R}(\chi_B) > 0$, then $\overline{R}(X|B) = \mathrm{GBR}(X|B)$.

- Naturally recover the generalized Bayes rule.
- Use this to define independence.

Related work

Standing on the shoulder of giants



- Gorban: empirical motivation. [1]
- Von Mises: strictly frequentist approach, "collectives". [3]
- Ivanenko: decision-theoretic approach to "non-stochastic" sequences (+ nets). [4]
- Walley: coherent upper previsions. [5]
- Walley & Fine: frequentist account of upper probabilities. [6]
  - In addition: long line of work by Fine and collaborators.

Future directions

What to do in practice?

- **Need randomness assumptions!**
- Introduce selection rules (à la von Mises) in a strictly frequentist way?
- Or pursue the approach of Fierens, Rêgo and Fine? [7]
- Connections to game-theoretic probability?

References

[1] Igor I. Gorban. The Statistical Stability Phenomenon. Springer, 2017.

[2] Andrei Nikolaevich Kolmogorov. On logical foundations of probability theory. In K. Ito and J.V. Prokhorov, editors, Probability theory and mathematical statistics: Proceedings of the Fourth USSR – Japan Symposium, 1982, pages 1—5. Springer, 1983.

[3] Richard von Mises and Hilda Geiringer. Mathematical theory of probability and statistics. Academic Press, 1964.

[4] Victor I. Ivanenko. Decision Systems and Nonstochastic Randomness. Springer, 2010.

[5] Peter Walley. Statistical reasoning with imprecise probabilities. Chapman-Hall, 1991.

[6] Peter Walley and Terrence L. Fine. Towards a frequentist theory of upper and lower probability. The Annals of Statistics, 10(3):741–761, 1982.

[7] Pablo Ignacio Fierens, Leonardo Chaves Rêgo, and Terrence L. Fine. A frequentist understanding of sets of measures. Journal of Statistical Planning and Inference, 139(6):1879–1892, 2009.