

# Towards a Strictly Frequentist Theory of Imprecise Probability

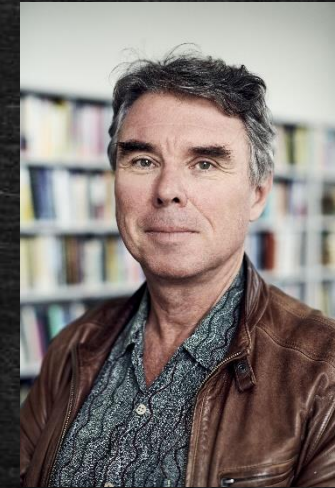
---

Christian Fröhlich, Rabanus Derr, Robert C. Williamson

# The “Foundations of Machine Learning Systems” Group

<http://fm.ls>

University of Tübingen; Tübingen AI Center



Led by  
Robert C.  
Williamson

# Core research interest: better models for data in ML

---

## *Data is Imperfect*

- Corruptions
- Missing data
- Dataset shift
- Data about people
  - Ethical concerns such as fairness!
  - Society in non-equilibrium

# Core research interest: better models for data in ML

---

## *Data is Imperfect*

- Corruptions
- Missing data
- Dataset shift
- Data about people
  - Ethical concerns such as fairness!
  - Society in non-equilibrium

## The standard i.i.d. assumption

- The utopia of *independent and identically distributed* data
- Just one representative example:

We consider supervised learning with a *risk-averse learner*. The learner has a data set comprised of **i.i.d. samples from an unknown distribution**, i.e.,  $D = \{(x_1, y_1), \dots, (x_N, y_N)\} \in (\mathcal{X} \times \mathcal{Y})^N \sim \mathcal{D}^N$ , and her goal is to learn a function  $h_\theta : \mathcal{X} \rightarrow \mathcal{R}$  that is parametrized by  $\theta \in \Theta \subset \mathbb{R}^d$ . The

[Curi et al., 2020]

The conceptual culprit:  
The law of large numbers

---

# The conceptual culprit: The law of large numbers

---

- Let  $X_1, X_2, \dots$  be i.i.d. random variables with finite expectation  $E[X]$ , then the sample average converges almost surely to the expectation

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = E[X]\right) = 1$$

# The conceptual culprit: The law of large numbers

---

- Let  $X_1, X_2, \dots$  be i.i.d. random variables with finite expectation  $E[X]$ , then the sample average converges almost surely to the expectation

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = E[X]\right) = 1$$

- Theoretical criticism: holds on a set of sequences with Lebesgue measure 1..not for *all* sequences.
  - Different ways of measuring the size of this set yield different answers!

# The conceptual culprit: The law of large numbers

---

- Let  $X_1, X_2, \dots$  be i.i.d. random variables with finite expectation  $E[X]$ , then the sample average converges almost surely to the expectation

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = E[X]\right) = 1$$

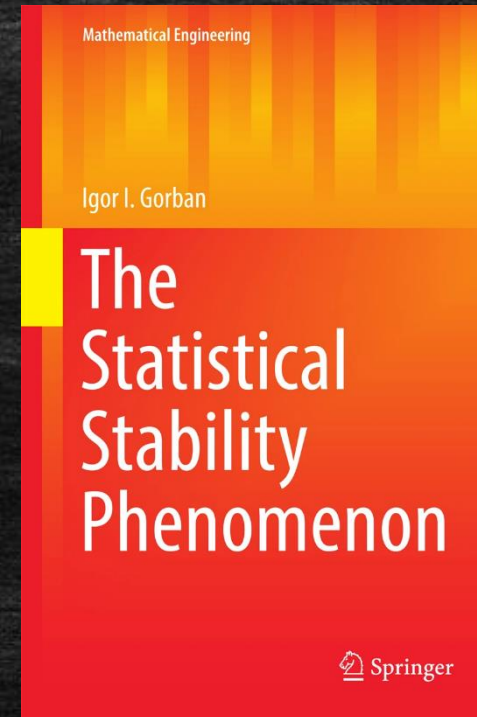
- Theoretical criticism: holds on a set of sequences with Lebesgue measure 1..not for *all* sequences.
  - Different ways of measuring the size of this set yield different answers!
- Serves to justify the *hypothesis of statistical stability* [Gorban, 2017]: that relative frequencies tend to stabilize in the long run



# Challenges to statistical stability

---

- From machine learning: dataset shift etc.
- Many interesting examples for instability [Gorban, 2017]
  - Mains voltage in a city
  - Earth's magnetic field
  - ...

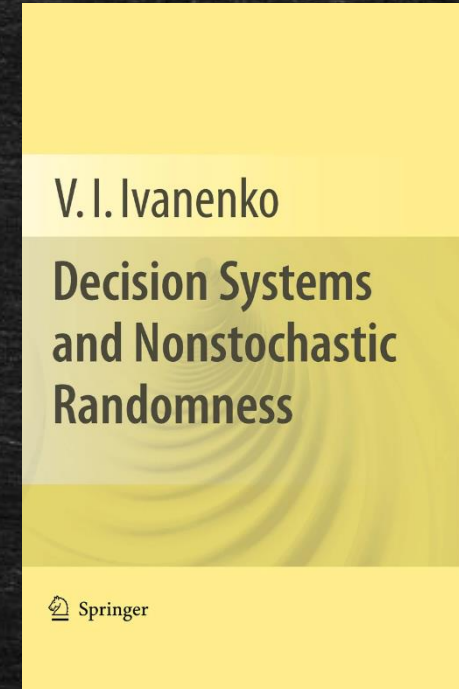
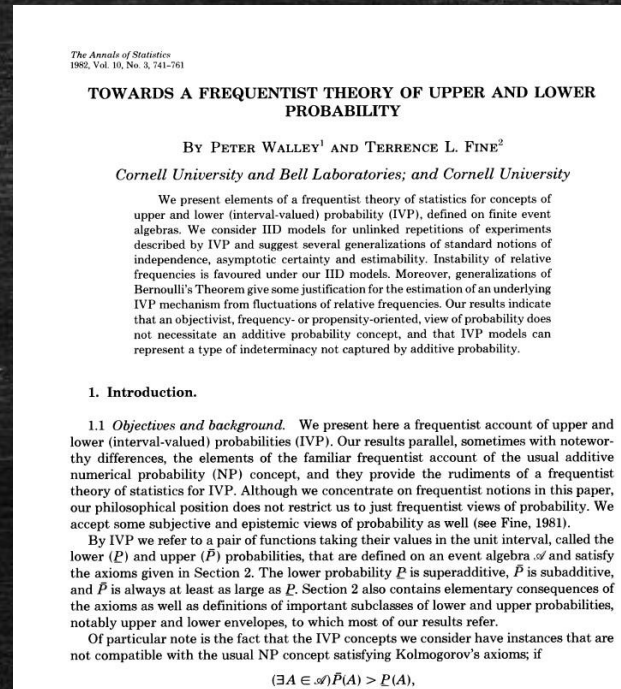
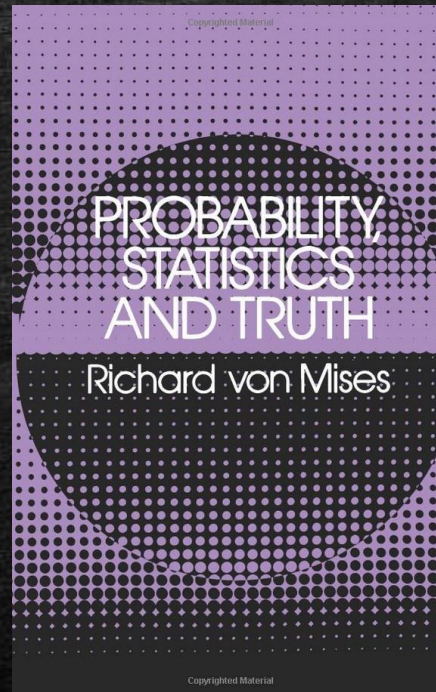


Our aim:

---

Find a principled way to model statistically unstable phenomena

How:



Tying together threads from von Mises, Walley & Fine, and Ivanenko

# A strictly frequentist approach

---

# A strictly frequentist approach

---

- Take the notion of a sequence as the primitive
  - In the spirit of von Mises and Ivanenko
  - Start with the data!

# A strictly frequentist approach

---

- Take the notion of a sequence as the primitive
  - In the spirit of von Mises and Ivanenko
  - Start with the data!
- Assume an arbitrary possibility set  $\Omega$  and  $L^\infty = \{X: \Omega \rightarrow \mathbb{R}: \sup_{\omega \in \Omega} |X(\omega)| < \infty\}$

# A strictly frequentist approach

---

- Take the notion of a sequence as the primitive
  - In the spirit of von Mises and Ivanenko
  - Start with the data!
- Assume an arbitrary possibility set  $\Omega$  and  $L^\infty = \{X: \Omega \rightarrow \mathbb{R}: \sup_{\omega \in \Omega} |X(\omega)| < \infty\}$
- Assume a sequence  $\Omega^\rightarrow: \mathbb{N} \rightarrow \Omega$  of elementary outcomes

# A strictly frequentist approach

---

- Take the notion of a sequence as the primitive
  - In the spirit of von Mises and Ivanenko
  - Start with the data!
- Assume an arbitrary possibility set  $\Omega$  and  $L^\infty = \{X: \Omega \rightarrow \mathbb{R}: \sup_{\omega \in \Omega} |X(\omega)| < \infty\}$
- Assume a sequence  $\Omega^\rightarrow: \mathbb{N} \rightarrow \Omega$  of elementary outcomes
- For any event  $A \subseteq \Omega$ , consider the limiting relative frequency:

$$P(A) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \chi_A(\Omega^\rightarrow(i))$$

But the limit may or may not exist! Relative frequencies could *diverge*.



Cluster points  $\rightarrow$  coherent upper prevision

---

# Cluster points -> coherent upper prevision

---

- Define a sequence of linear previsions as:

$$E^{\rightarrow}(n) := X \mapsto \frac{1}{n} \sum_{i=1}^n X(\Omega^{\rightarrow}(i)), \quad X \in L^{\infty}$$

# Cluster points $\rightarrow$ coherent upper prevision

---

- Define a sequence of linear previsions as:

$$E^{\rightarrow}(n) := X \mapsto \frac{1}{n} \sum_{i=1}^n X(\Omega^{\rightarrow}(i)), \quad X \in L^{\infty}$$

- From this, induce coherent upper prevision:

$$\bar{R}(X) := \sup\{E(X) : E \in \text{CP}(E^{\rightarrow})\}, \quad X \in L^{\infty}$$

- Where  $\text{CP}(\cdot)$  is the set of cluster points with respect to the weak\* topology on the dual of  $L^{\infty}$
- This set is guaranteed to be non-empty!

# Cluster points -> coherent upper prevision

---

- Define a sequence of linear previsions as:

$$E^{\rightarrow}(n) := X \mapsto \frac{1}{n} \sum_{i=1}^n X(\Omega^{\rightarrow}(i)), \quad X \in L^{\infty}$$

- From this, induce coherent upper prevision:

$$\bar{R}(X) := \sup\{E(X) : E \in \text{CP}(E^{\rightarrow})\}, \quad X \in L^{\infty}$$

- Where  $\text{CP}(\cdot)$  is the set of cluster points with respect to the weak\* topology on the dual of  $L^{\infty}$
- This set is guaranteed to be non-empty!

- Proposition:  $\bar{R}(X) = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X(\Omega^{\rightarrow}(i))$

Limit is to Expectation  
as  
Cluster Points to Upper Prevision

# Take-away:

---

Even if “the” probability does not exist, we always have a coherent upper prevision.  
Also: can recover the generalized Bayes rule and formulate independence.

# The converse direction

---

- **Theorem:** for finite  $\Omega$  and arbitrary coherent upper prevision  $\bar{R}$  on  $\Omega$ , can always construct a corresponding sequence  $\Omega^\rightarrow$  (such that  $\Omega^\rightarrow$  induces  $\bar{R}$ )
- Proof is constructive!

# The converse direction

---

- **Theorem:** for finite  $\Omega$  and arbitrary coherent upper prevision  $\bar{R}$  on  $\Omega$ , can always construct a corresponding sequence  $\Omega^\rightarrow$  (such that  $\Omega^\rightarrow$  induces  $\bar{R}$ )
- Proof is constructive!

=> Strictly frequentist re-interpretation of subjective upper previsions.

# Future Directions – What to do in practice?


---

## The finite data setting

- Need randomness assumptions
- Connections to game-theoretic randomness?
- How is this related to work of Fierens, Rêgo and Fine?
  - “A frequentist understanding of sets of measures”




# Come visit our poster ☺



**UNIVERSITÄT TÜBINGEN**

## Towards a Strictly Frequentist Theory of Imprecise Probability

Christian Fröhlich<sup>1</sup>, Rabanus Derr<sup>2</sup>, Robert C. Williamson<sup>1,2</sup>  
University of Tübingen, Tübingen AI Center




**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 equilibria

### The converse direction


From upper prevision to sequences

**Theorem:** Let  $\|\beta\| < \infty$ . Let  $\mathbb{P}$  be a coherent upper prevision on  $L^{\infty}$ . There exists a sequence  $\{\beta\}$  such that we can write  $\mathbb{P}$  as  $(\forall X \in L^{\infty})$

$$\mathbb{P}(X) = \sup\{E(X); E \in \mathcal{E}_{\beta}\}, \mathcal{E}_{\beta} \subseteq \mathcal{CP}(\mathcal{F}_{\beta})$$

where  $\mathcal{E}_{\beta}(\beta) := X \mapsto \sum_{i=1}^n \lambda_i(\beta(i)) \sum_{\omega \in \Omega} \lambda_{\omega}(\beta(i))$

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



- ▶ Wolley & Fine [5] offer similar result for upper probabilities

---

### New models for data in ML

An overarching agenda

**Data has many imperfections:**

- ▶ Data corruption
- ▶ Dataset shift: not a single, stable distribution
- ▶ need to move beyond the IID, assumption (independent and identically distributed)

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

### Conditional upper prevision

Inspired by von Mises

Sequence of conditional linear previsions  $(\lambda_n)$  is the indicator function of  $\mathcal{B}$ :

$$\mathbb{E}_i(\beta)(\omega) := X \mapsto \sum_{j=1}^n \lambda_n(X, \omega)(\beta(i)) \sum_{\omega \in \Omega} \lambda_{\omega}(\beta(i))$$

Define the **conditional upper prevision**

$$\mathbb{P}_i(X, \beta) := \sup\{E(X); E \in \mathcal{CP}(\mathcal{F}_i(\beta))\}, \forall X \in L^{\infty}$$

For  $\mathbb{P}_i(X, \beta) > 0$ , define the conditional set of desirable gambles as:

$$D_{\mathbb{P}_i} := \{X \in L^{\infty}; \mathbb{P}_i(X, \beta) \leq 0\}$$

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

$$\text{GBR}(X, \beta) := \inf\{v \in \mathbb{R}; X - v \in D_{\mathbb{P}_i}\}$$

**Proposition:** If  $\mathbb{P}_i(\beta) > 0$ , then  $\mathbb{P}_i(X, \beta) = \text{GBR}(X, \beta)$

- ▶ Naturally recovers the generalized Bayes rule
- ▶ Use this to define independence


---

### Strict frequentism + divergence

- ▶ Aim: principled way to model **unstable** data
- ▶ Strict frequentist **sequence** as the primitive – cf. von Mises collectives
- ▶ Where the “probability” does not exist, natural replacement:  $\beta$  on upper prevision
- ▶ Coherent **upper probability** as limiting relative frequency
- ▶ In their latest paper of coherent previsions and coherent upper previsions
- ▶ Works for all sequences and all events
- ▶ Generally, 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

### Related work

Standing on the shoulders of giants



- ▶ Gärdenfors: empirical motivation [1]
- ▶ von Mises: strictly frequentist approach, “collectives” [2]
- ▶ Bayesian, decision-theoretic approach to “non-stochastic” sequences (+ notes) [4]
- ▶ Wolley: coherent upper previsions [3]
- ▶ Wolley & Fine: frequentist account of upper probabilities [5]
- ▶ In addition: long list of work by Fine and collaborators

---

### Technical Setup

Our setting

- ▶ Let  $\Omega$  be an arbitrary set of outcomes. A gamble is a bounded function  $X: \Omega \rightarrow \mathbb{R}$
- ▶ We assume a loss-based orientation, + is loss, - is gain
- ▶ The set  $L^{\infty}$  of bounded gambles forms a Banach space, with dual  $(L^{\infty})^*$  on which we install the weak\* topology on  $(L^{\infty})^*$
- ▶ The set of linear previsions is compact under the weak\* topology
- ▶  $\mathcal{CP}(\mathcal{F}) := \{E \in (L^{\infty})^*; E(X) \geq 0 \forall X \in \mathcal{C}, E(\mathbb{1}_{\Omega}) = 1\} \subset (L^{\infty})^*$
- ▶ Let  $\mathbb{P}: N \rightarrow [0, 1]$ , an  $\mathcal{O}$ -valued sequence of elementary outcomes
- ▶ Induces sequence of linear previsions:  $\mathbb{E}(\beta) := X \mapsto \sum_{i=1}^n \lambda_i(\beta(i)) \sum_{\omega \in \Omega} \lambda_{\omega}(\beta(i))$

### 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 Fine, Fine and Fine [7]
- ▶ Connections to game-theoretic probability?

---

### The forward direction

Allowing for divergence

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

$$\mathbb{P}(X) := \sup\{E(X); E \in \mathcal{CP}(\mathcal{F}_{\beta})\}, \forall X \in L^{\infty}, \mathbb{E}(X) := \mathbb{P}(X)$$

**Proposition:**

$$\mathcal{CP}\left(\beta := \frac{1}{n} \sum_{i=1}^n \lambda_i(\beta(i))\right) = \{E(X); E \in \mathcal{CP}(\mathcal{F}_{\beta})\}, \forall X \in L^{\infty}$$

Therefore:

$$\mathbb{P}(X) = \inf_{\beta} \sup_{E \in \mathcal{CP}(\mathcal{F}_{\beta})} E(X)$$

- ▶ When the limiting relative frequency does not exist, use the **lim sup** for decision making!

### References

[1] Iain V. Gärdenfors, The Statistical Stability Postulate, Springer, 2017  
[2] Felix Richard Hahn, The Logic of Probabilities, in: K. B. E. and J. V. (eds.), Probability, Frequentist Theory and Mathematical Models, Proceedings of the Fourth USSR - Soviet Symposium, 1982, pages 1-6, Springer, 1982  
[3] Robert von Mises and Hans Lange, Mathematical Theory of Probability and Statistics, Academic Press, 1980  
[4] Robert von Mises, Decision Theory and Bayesian Decision, Springer, 2000  
[5] Peter Wolley, Coherent previsions and non-stochastic sequences, arXiv:1801.08811, 2018  
[6] Peter Wolley, Coherent previsions and non-stochastic sequences, arXiv:1801.08811, 2018  
[7] Peter Wolley and Robert C. Williamson, The Theory of Coherent Previsions of Upper and Lower Probabilities, The Journal of Statistics, 2017, 16, 362  
[8] Peter Wolley, Robert C. Williamson, Thomas Rippl, and Robert C. Williamson, A frequentist understanding of sets of measures, Journal of Probability Theory and Statistics, 2019, 2019, 2019

ISITPA 2023 <http://fs.18>

Happy to chat about machine learning and IP

13

