Measurability of Definable Sets over Tame Ordered Fields

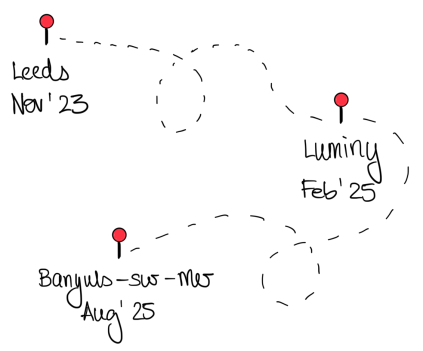
(joint work with L. S. Krapp and M. Vermeil: arXiv:2506.08733)

Laura Wirth
University of Konstanz

DDG40: Structures algébriques ordonnées
Banyuls-sur-Mer

Question.

Given a subfield $K \subseteq \mathbb{R}$, is every definable set $A \subseteq K^n$ Borel measurable?



Central Question – Partial Answers

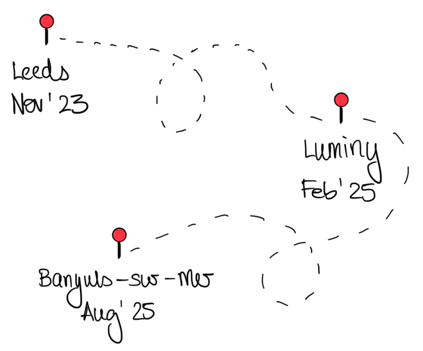
Question.

Given a subfield $K \subseteq \mathbb{R}$, is every definable set $A \subseteq K^n$ Borel measurable?

Partial Answers.

Yes, if

- K is real closed,
- K is countable.



Context

Context: Central Objects

Model Theory

Statistical Learning Theory

structure $\mathcal{M}$

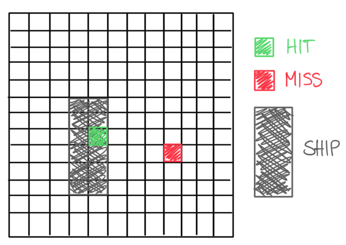
hypothesis space $\mathcal{H}$

Hypothesis Spaces

Learning Framework.

- instance space $\emptyset \neq \mathcal{X}$
- sample space $\mathcal{Z} = \mathcal{X} \times \{0, 1\}$
- hypothesis space $\emptyset \neq \mathcal{H} \subseteq \{0, 1\}^{\mathcal{X}}$

Example: Battleship.



Definition.

Let $\mathcal{L}$ be a language, let $\mathcal{M}$ be an $\mathcal{L}$–structure and let $\varphi(x_1, \ldots, x_n; p_1, \ldots, p_\ell)$ be an $\mathcal{L}$–formula. For any $\boldsymbol{w} \in M^\ell$, set

$$\varphi(\mathcal{M}; \boldsymbol{w}) = \{\boldsymbol{a} \in M^n \mid \mathcal{M} \models \varphi(\boldsymbol{a}; \boldsymbol{w})\}.$$

Then the hypothesis space $\mathcal{H}^\varphi \subseteq \{0,1\}^{M^n}$ is given by

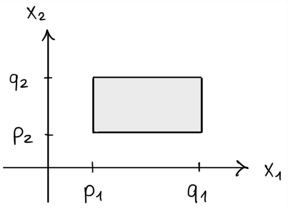
$$\mathcal{H}^\varphi := \left\{ \mathbb{1}_{\varphi(\mathcal{M};\boldsymbol{w})} \,\middle|\, \boldsymbol{w} \in M^\ell \right\}.$$

Definable Hypothesis Spaces: Example

Set $\mathcal{L}_{\text{or}} := \{+, \cdot, -, 0, 1, <\}$, $\mathbb{R}_{\text{or}} := (\mathbb{R}, +, \cdot, -, 0, 1, <)$ and consider the $\mathcal{L}$–formula $\varphi(x_1, x_2; p_1, q_1, p_2, q_2)$ given by

$$p_1 \leq x_1 \leq q_1 \wedge p_2 \leq x_2 \leq q_2.$$

For $w = (p_1, q_1, p_2, q_2) \in \mathbb{R}^4$, the set $\varphi(\mathbb{R}_{\text{or}}; w)$ is an axis-aligned rectangle in $\mathbb{R}^2$ of the form:

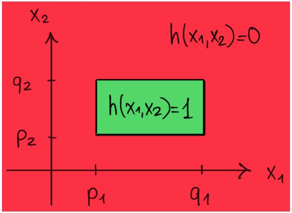


Definable Hypothesis Spaces: Example

Set $\mathcal{L}_{\mathrm{or}} := \{+, \cdot, -, 0, 1, <\}$, $\mathbb{R}_{\mathrm{or}} := (\mathbb{R}, +, \cdot, -, 0, 1, <)$ and consider the $\mathcal{L}$–formula $\varphi(x_1, x_2; p_1, q_1, p_2, q_2)$ given by

$$p_1 \leq x_1 \leq q_1 \wedge p_2 \leq x_2 \leq q_2.$$

For $w = (p_1, q_1, p_2, q_2) \in \mathbb{R}^4$, the set $\varphi(\mathbb{R}_{\mathrm{or}}; w)$ is an axis-aligned rectangle in $\mathbb{R}^2$ of the form:



The hypothesis $h = \mathbb{1}_{\varphi(\mathbb{R}_{\mathrm{or}}; w)} \in \mathcal{H}^{\varphi}$ sends all points inside this rectangle to 1 and all points outside to 0.

Context: Central Notions

Model Theory

- o-Minimality
- structure $\mathcal{M}$
- NIP

Statistical Learning Theory

- PAC Learning
- hypothesis space $\mathcal{H}$
- VC Dimension

The following notion is due to Pillay and Steinhorn 1986.

Recall.

Given a language $\mathcal{L} = \{<, \dots\}$ and an $\mathcal{L}$–structure $\mathcal{M} = (M, <, \dots)$ for which $(M, <)$ is a linear order, $\mathcal{M}$ is called o-minimal if for any $\mathcal{L}$–formula $\varphi(x; p_1, \dots, p_\ell)$ and any $\boldsymbol{w} \in M^\ell$ the set $\varphi(\mathcal{M}; \boldsymbol{w}) \subseteq M$ can be expressed as a finite union of points and open intervals.

They further related it to the notion NIP, which was introduced by Shelah 1971.

Proposition.

If $\mathcal{M} = (M, <, \dots)$ is o-minimal, then $\mathcal{M}$ has NIP.

A. Pillay and C. Steinhorn, 'Definable sets in ordered structures', I, *Trans. Amer. Math. Soc.* **295** (1986) 565–592.

S. Shelah, 'Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory', *Ann. Math. Logic* **3** (1971) 271–362.

The following result is due to Laskowski 1992.

Proposition.

Let $\mathcal{L}$ be a language and let $\mathcal{M}$ be an $\mathcal{L}$–structure. Then the following conditions are equivalent:

(1) $\mathcal{M}$ has NIP.

(2) The hypothesis space $\mathcal{H}^\varphi$ has finite VC dimension for any $\mathcal{L}$–formula $\varphi(x; p)$.

M. C. Laskowski, 'Vapnik–Chervonenkis classes of definable sets', *J. Lond. Math. Soc.* **45** (1992) 377–384.

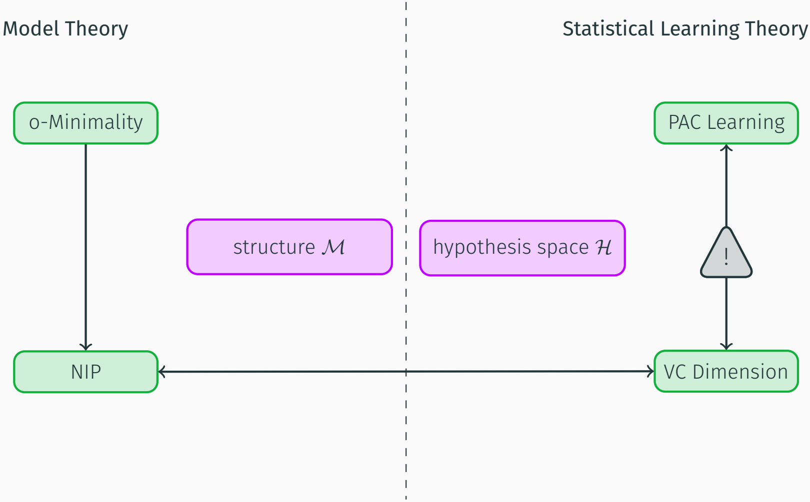
Fundamental Theorem of Statistical Learning.
Let $\mathcal{H}$ be well-behaved. Then $\mathcal{H}$ is PAC learnable if and only if $\mathcal{H}$ has finite VC dimension.

Originally, this equivalence result is due to Blumer, Ehrenfeucht, Haussler and Warmuth 1989.

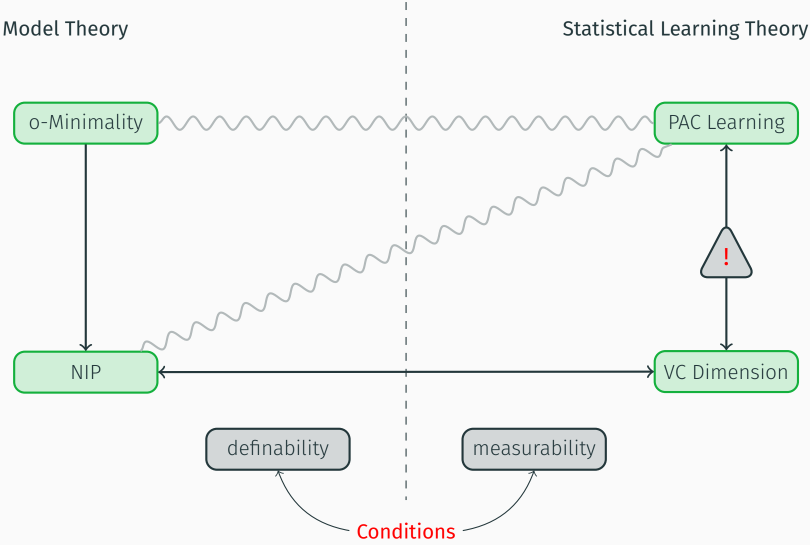
A. Blumer, A. Ehrenfeucht, D. Haussler and M. K. Warmuth, 'Learnability and the Vapnik-Chervonenkis dimension', *J. Assoc. Comput. Mach.* **36** (1989) 929–965.

L. S. Krapp and L. Wirth, 'Measurability in the Fundamental Theorem of Statistical Learning', Preprint, 2025, arXiv:2410.10243.

Context: Relationship of Central Conditions



Context: Central Conditions



Measurability

Recall.

A probability space $(\Omega, \Sigma, \mathbb{P})$ consists of

- a domain Ω,
- a σ–algebra $\Sigma \subseteq \mathcal{P}(\Omega)$ (containing the measurable subsets),
- and a probability measure $\mathbb{P} \colon \Sigma \to [0, 1]$ (also called distribution).

Given a probability space $(\Omega, \Sigma, \mathbb{P})$, a map $g \colon \Omega \to \mathbb{R}$ is called Σ–measurable if $g^{-1}(B) \in \Sigma$ for any Borel set $B \subseteq \mathbb{R}$.

Fix a σ–algebra $\Sigma_{\mathcal{Z}}$ on the sample space $\mathcal{Z} = \mathcal{X} \times \{0, 1\}$.

Error and Sample Error

Let $h \in \mathcal{H}$ be a hypothesis, let $\mathbb{D}$ be a distribution on $(\mathcal{Z}, \Sigma_{\mathcal{Z}})$, and let $\boldsymbol{z} = (z_1, \ldots, z_m) \in \mathcal{Z}^m$ be a multi-sample.

The (true) error of h is given by

$$\mathrm{er}_{\mathbb{D}}(h) := \mathbb{D}(\{(x, y) \in \mathcal{Z} \mid h(x) \neq y\}) = \mathbb{D}(\mathcal{Z} \setminus \Gamma(h)).$$

The sample error of h on $\boldsymbol{z} = (z_1, \ldots, z_m) \in \mathcal{Z}^m$, given by

$$\hat{\mathrm{er}}_{\boldsymbol{z}}(h) := \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}_{\mathcal{Z} \setminus \Gamma(h)}(z_i),$$

provides a useful estimate for the true error.

S. Ben-David and S. Shalev-Shwartz, *Understanding Machine Learning: From Theory to Algorithms*, (Cambridge University Press, Cambridge, 2014).

Remarks.

- The error $\mathrm{er}_{\mathbb{D}}(h) = \mathbb{D}(\mathcal{Z} \setminus \Gamma(h))$ is well-defined iff $\Gamma(h) \in \Sigma_{\mathcal{Z}}$.
- If $\Gamma(h) \in \Sigma_{\mathcal{Z}}$, then the map

$$\begin{aligned} \mathcal{Z}^m &\to \left\{ \tfrac{k}{m} \;\middle|\; k \in \{0, 1, \ldots, m\} \right\}, \\ \mathbf{z} &\mapsto \hat{\mathrm{er}}_{\mathbf{z}}(h) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}_{\mathcal{Z} \setminus \Gamma(h)}(z_i) \end{aligned}$$

  is $\Sigma_{\mathcal{Z}}^m$–measurable.

Measurability of Definable Sets

Borel σ–algebra

Let $(K, <)$ be an ordered field and let $n \in \mathbb{N}$.

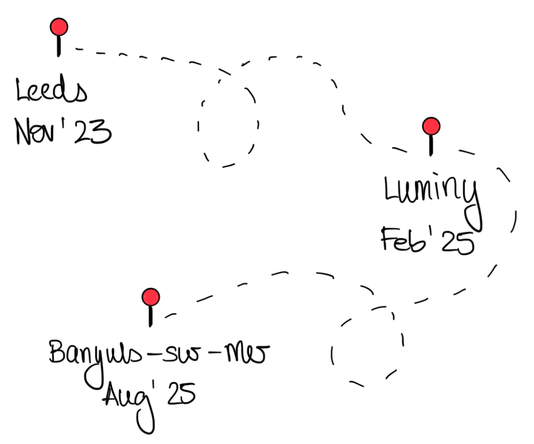
Endow K with the order topology τ, K^n with the product topology τ^n, and denote by $\mathcal{B}(K^n)$ the Borel σ–algebra on K^n generated by τ^n.

Recall that $\mathcal{B}(K^n)$ is the smallest σ–algebra containing τ^n as a subset.

For $X \subseteq K^n$, consider the trace σ–algebra given by

$$\mathcal{B}(X) := \{B \cap X \mid B \in \mathcal{B}(K^n)\}.$$

Towards an Answer to the Central Question



Defining a Non-Borel Set

Theorem.

There exists a subfield $K \subseteq \mathbb{R}$ that has the independence property (i.e. it is not NIP) and defines a set $D \subseteq K$ with $D \notin \mathcal{B}(K)$.

Question.

Given a subfield $K \subseteq \mathbb{R}$, is every definable set $A \subseteq K^n$ Borel measurable?

$\longrightarrow$ Generally, no.

L. S. Krapp, M. Vermeil and L. Wirth, 'On Tameness, Measurability and the Independence Property', Preprint, 2025, arXiv:2506.08733.

Proof Sketch

- By a technical construction there exist two disjoint *measure-irregular* sets $A, A' \subseteq \mathbb{R}$ (of cardinality $\mathfrak{c}$) such that $A \dot\cup A'$ is algebraically independent over $\mathbb{Q}$.
- Set $\sqrt{A_{\geq 0}} = \{\sqrt{a} \mid a \in A, a \geq 0\}$ and $K = \mathbb{Q}\left(\sqrt{A_{\geq 0}} \cup A'\right)$.
- Write $K = F(t)$, where F is a subfield of K and t is transcendental over F. By R. Robinson 1964, K defines $\mathbb{Z}$. Therefore, K has the independence property.
- Consider the formula $\exists y\, x = y^2$ defining the set $D = \{y^2 \mid y \in K\} \subseteq K$.
- Note that $A_{\geq 0} \subseteq D$ and $A' \cap D = \emptyset$.
- Derive from the *measure-irregularity* of A and A' that $D \notin \mathcal{B}(K)$.

□

R. M. Robinson, 'The undecidability of pure transcendental extensions of real fields', *Z. Math. Logik Grundlagen Math.* **10** (1964) 275–282.

Remarks

The field K is a purely transcendental extension of $\mathbb{Q}$ by continuum many algebraically independent elements.

Wild properties of the (ordered) field K:

- not o-minimal
- not real closed
- not almost real closed
- undecidable
- admits $2^{\mathfrak{c}}$ many non-isomorphic archimedean orderings and $2^{\mathfrak{c}}$ many non-isomorphic non-archimedean orderings

Lemma.

Let $\mathcal{L}$ be a language expanding $\mathcal{L}_{\mathrm{or}}$, let $\mathcal{R}$ be an o-minimal $\mathcal{L}$–expansion of an ordered field, let $n \in \mathbb{N}$ and let $A \subseteq R^n$ be $\mathcal{L}$–definable. Then $A \in \mathcal{B}(R^n)$.

Finish.

M. Karpinski and A. Macintyre, 'Approximating Volumes and Integrals in o-Minimal and p-Minimal Theories', *Connections between model theory and algebraic and analytic geometry* (ed. Macintyre), Quad. Mat. 6 (2000) 149–177.

Guided by the work of Karpinski and Macintyre 2000 we proved the following learnability result:

Theorem. Let

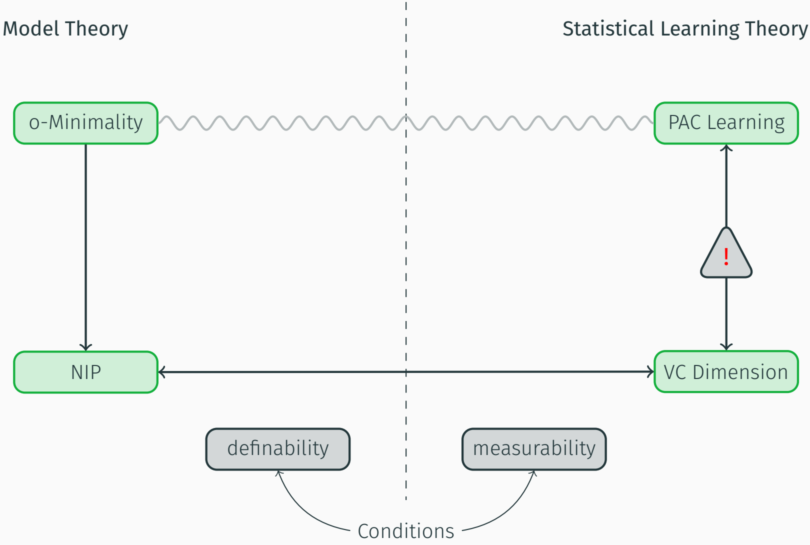
- $\mathcal{L}$ be a language expanding $\mathcal{L}_{\mathrm{or}}$,
- $\mathcal{R}$ be an o-minimal $\mathcal{L}$–expansion of $\mathbb{R}_{\mathrm{or}}$,
- $\varphi(x_1, \ldots, x_n; p_1, \ldots, p_\ell)$ be an $\mathcal{L}$–formula,
- $\Sigma_{\mathcal{Z}}$ be a σ–algebra on $\mathcal{Z} = \mathbb{R}^n \times \{0, 1\}$ with $\mathcal{B}(\mathcal{Z}) \subseteq \Sigma_{\mathcal{Z}}$, and
- $\mathcal{D}$ be a set of distributions on $(\mathcal{Z}, \Sigma_{\mathcal{Z}})$ such that $(\mathcal{Z}^m, \Sigma_{\mathcal{Z}}^m, \mathbb{D}^m)$ is a complete probability space for any $\mathbb{D} \in \mathcal{D}$ and any $m \in \mathbb{N}$.

Then $\mathcal{H}^\varphi$ is PAC learnable with respect to $\mathcal{D}$.

Skip proof.

L. S. Krapp and L. Wirth, 'Measurability in the Fundamental Theorem of Statistical Learning', Preprint, 2025, arXiv:2410.10243.

Roadmap


Model Theory
Statistical Learning Theory
o-Minimality
PAC Learning
!
NIP
VC Dimension
definability
measurability
Conditions

Relevant Functions

Definition.

A hypothesis space $\mathcal{H}$ is called well-behaved with respect to the set $\mathcal{D}$ of distributions on $(\mathcal{Z}, \Sigma_{\mathcal{Z}})$ if it satisfies the following conditions:

- $\Gamma(h) \in \Sigma_{\mathcal{Z}}$ for any $h \in \mathcal{H}$.
- There exists $m_{\mathcal{H}} \in \mathbb{N}$ such that the map

$$U\colon \mathcal{Z}^m \to [0,1],\ \mathbf{z} \mapsto \sup_{h \in \mathcal{H}} \left| \mathrm{er}_{\mathbb{D}}(h) - \hat{\mathrm{er}}_{\mathbf{z}}(h) \right|$$

  is $\Sigma_{\mathcal{Z}}^m$–measurable for any $m \geq m_{\mathcal{H}}$ and any $\mathbb{D} \in \mathcal{D}$, and the map

$$V\colon \mathcal{Z}^{2m} \to [0,1],\ (\mathbf{z}, \mathbf{z}') \mapsto \sup_{h \in \mathcal{H}} \left| \hat{\mathrm{er}}_{\mathbf{z}'}(h) - \hat{\mathrm{er}}_{\mathbf{z}}(h) \right|$$

  is $\Sigma_{\mathcal{Z}}^{2m}$–measurable for any $m \geq m_{\mathcal{H}}$.

Proof Sketch.

- o-Minimality implies NIP.
- Thus, $\mathcal{H}^{\varphi}$ has finite VC dimension.
- Aim: Apply Fundamental Theorem.
- To this end: Verify that $\mathcal{H}^{\varphi}$ is well-behaved.
- $\Gamma(h) \in \mathcal{B}(\mathcal{Z}) \subseteq \Sigma_{\mathcal{Z}}$ for any $h \in \mathcal{H}^{\varphi}_{\mathcal{X}}$ (follows from Lemma).
- Technical analysis and application of Pollard's arguments regarding measurability of suprema establish measurability of the maps U and V.

$\square$

D. Pollard, *Convergence of Stochastic Processes*, Springer Ser. Stat. (Springer, New York, 1984).

Capacity for Enhancement of the Learnability Theorem

- Can the result be extended to further tame (e.g. o-minimal) structures?

  $\longrightarrow$ Yes, but the topological space should satisfy the following properties that are crucial in the proof: metrizability, separability and local compactness

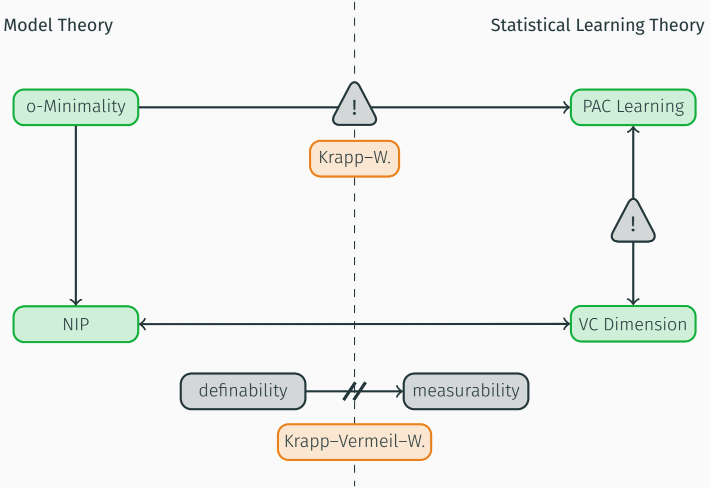
- Can the o-minimality assumption be weakended (e.g. to NIP)?

  $\longrightarrow$ This leads to the following open question:

Question.

Let K be an NIP ordered field and let $D \subseteq K$ be definable in K.
Is D necessarily Borel, i.e. $D \in \mathcal{B}(K)$?

Summary



L. S. Krapp and L. Wirth, 'Measurability in the Fundamental Theorem of Statistical Learning', Preprint, 2025, arXiv:2410.10243.
L. S. Krapp, M. Vermeil and L. Wirth, 'On Tameness, Measurability and the Independence Property', Preprint, 2025, arXiv:2506.08733.

Appendix

Learning over o-Minimal Expansions of $\mathbb{R}_{\mathrm{or}}$ – Remarks

Definition.

A probability space $(\Omega, \Sigma, \mathbb{P})$ is called complete if

$$\forall N \in \Sigma \; (\mathbb{P}(N) = 0 \Rightarrow \forall A \subseteq N \; A \in \Sigma).$$

Remarks.

- Completeness condition is crucial for deducing measurability of the maps U and V.
- Completeness condition is trivially satisfied if $\mathcal{Z}$ is countable, since then $\Sigma_{\mathcal{Z}} = \mathcal{B}(\mathcal{Z}) = \mathcal{P}(\mathcal{Z})$.
- Potential Solution: Replace product spaces with their respective completions at all relevant places.

  $\longrightarrow$ All impacted definitions and proofs need to be adjusted accordingly!

Jump back.

Further Remarks.

- The measurability of V can be established without imposing further conditions (like e.g. the completeness condition), since it can be shown to be definable.
  $\longrightarrow$ Unfortunately, this approach does not work for U.
- Further Work: Extend theorem to general o-minimal structures.

Jump back.
Jump to Summary.

Sufficient Conditions for Well-Behavedness

Remark.

Sufficient conditions for the measurability of the maps U and V:

- $\mathcal{X}$ resp. $\mathcal{Z}$ is countable.
- $\mathcal{H}$ is countable.
- $\mathcal{H}$ is universally separable.

Definition.

The hypothesis space $\mathcal{H}$ is called universally separable if there exists a countable subset $\mathcal{H}_0 \subseteq \mathcal{H}$ such that for any $h \in \mathcal{H}$ there exists a sequence $\{h_n\}_{n\in\mathbb{N}} \subseteq \mathcal{H}_0$ converging pointwise to h.

Notation.

$[m] := \{1, \ldots, m\}$ for $m \in \mathbb{N}$.

Definition.

Let $\mathcal{L}$ be a language and let $\mathcal{M}$ be an $\mathcal{L}$–structure.
A (partitioned) $\mathcal{L}$–formula $\varphi(x_1, \ldots, x_n; p_1, \ldots, p_\ell)$ has NIP over $\mathcal{M}$ if there is $m \in \mathbb{N}$ such that for any object set $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\} \subseteq M^n$ and any parameter set $\{\boldsymbol{w}_I \mid I \subseteq [m]\} \subseteq M^\ell$, there is some $J \subseteq [m]$ such that

$$\mathcal{M} \not\models \underbrace{\bigwedge_{i \in J} \varphi(\boldsymbol{a}_i; \boldsymbol{w}_J) \wedge \bigwedge_{i \in [m] \setminus J} \neg\varphi(\boldsymbol{a}_i; \boldsymbol{w}_J)}_{\varphi(\boldsymbol{a}_i; \boldsymbol{w}_J) \text{ is true iff } i \in J}.$$

S. Shelah, 'Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory', *Ann. Math. Logic* **3** (1971) 271–362.

Definition.

Let $\mathcal{L}$ be a language and let $\mathcal{M}$ be an $\mathcal{L}$–structure.
A (partitioned) $\mathcal{L}$–formula $\varphi(x_1, \ldots, x_n; p_1, \ldots, p_\ell)$ has NIP over $\mathcal{M}$ if there is $m \in \mathbb{N}$ such that for any object set $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\} \subseteq M^n$ and any parameter set $\{\boldsymbol{w}_I \mid I \subseteq [m]\} \subseteq M^\ell$, there is some $J \subseteq [m]$ such that

$$\mathcal{M} \not\models \underbrace{\bigwedge_{i \in J} \varphi(\boldsymbol{a}_i; \boldsymbol{w}_J) \wedge \bigwedge_{i \in [m] \setminus J} \neg\varphi(\boldsymbol{a}_i; \boldsymbol{w}_J)}_{\varphi(\boldsymbol{a}_i; \boldsymbol{w}_J) \text{ is true iff } i \in J}.$$

The $\mathcal{L}$–structure $\mathcal{M}$ has NIP if every $\mathcal{L}$–formula has NIP over $\mathcal{M}$.

Jump back.

S. Shelah, ‘Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory’, *Ann. Math. Logic* **3** (1971) 271–362.

Definition.

Given a hypothesis space $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ and a set $A \subseteq \mathcal{X}$, we say that $\mathcal{H}$ shatters A if

$$\{h\restriction_A \mid h \in \mathcal{H}\} = \{0,1\}^A.$$

If $\mathcal{H}$ cannot shatter sets of arbitrarily large size, then we say that $\mathcal{H}$ has finite VC dimension. In this case:

$$\mathrm{vc}(\mathcal{H}) := \max\{d \in \mathbb{N} \mid \exists A \subseteq \mathcal{X}, |A| = d \colon \mathcal{H} \text{ shatters } A\}.$$

Jump back.

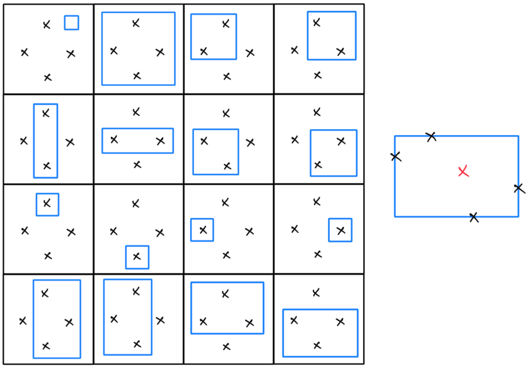
V. N. Vapnik and A. Ja. Červonenkis, 'Uniform Convergence of Frequencies of Occurrence of Events to Their Probabilities', *Dokl. Akad. Nauk SSSR* **181** (1968) 781–783 (Russian), *Sov. Math. Dokl.* **9** (1968) 915–918 (English).

Example

Consider $\mathbb{R}_{\text{or}}$ and the hypothesis space $\mathcal{H}^{\varphi}$ defined by the $\mathcal{L}_{\text{or}}$–formula $\varphi(x_1, x_2; p_1, q_1, p_2, q_2)$ given by

$$p_1 \leq x_1 \leq q_1 \wedge p_2 \leq x_2 \leq q_2.$$

We compute $\text{vc}(\mathcal{H}^{\varphi}) = 4$:



Jump back.

NIP Formulas and VC Dimension

The following result is due to Laskowski 1992.

Lemma.

Let $\mathcal{L}$ be a language, let $\mathcal{M}$ be an $\mathcal{L}$–structure and let $\varphi(x_1, \ldots, x_n; p_1, \ldots, p_\ell)$ be an $\mathcal{L}$–formula. Then φ has NIP over $\mathcal{M}$ if and only if the hypothesis space $\mathcal{H}^\varphi$ has finite VC dimension.

M. C. Laskowski, 'Vapnik–Chervonenkis classes of definable sets', *J. Lond. Math. Soc.* **45** (1992) 377–384.

Learning from Examples: Mathematically

A learning function is a map of the form $\mathcal{A}\colon \bigcup_{m \in \mathbb{N}} \mathcal{Z}^m \to \mathcal{H}$.

The input for $\mathcal{A}$ is generated according to an arbitrary distribution $\mathbb{D} \in \mathcal{D}$:

$$\boldsymbol{z} = \underbrace{((x_1, y_1), \ldots, (x_m, y_m))}_{\text{iid samples } \sim \mathbb{D}^m} \in \mathcal{Z}^m.$$

$\mathcal{A}$ then predicts a generalization hypothesis $h = \mathcal{A}(\boldsymbol{z}) \in \mathcal{H}$ based on the multi-sample $\boldsymbol{z}$.

S. Ben-David and S. Shalev-Shwartz, *Understanding Machine Learning: From Theory to Algorithms*, (Cambridge University Press, Cambridge, 2014).

Learning from Examples – Goal

The goal is to minimize the (true) error of h given by

$$\mathrm{er}_{\mathbb{D}}(h) := \mathbb{D}(\{(x,y) \in \mathcal{Z} \mid h(x) \neq y\}) = \mathbb{D}(\underbrace{\mathcal{Z} \setminus \Gamma(h)}_{\in \Sigma_{\mathcal{Z}}}).$$

More precisely, we want to achieve an error that is close to

$$\mathrm{opt}_{\mathbb{D}}(\mathcal{H}) := \inf_{h \in \mathcal{H}} \mathrm{er}_{\mathbb{D}}(h).$$

S. Ben-David and S. Shalev-Shwartz, *Understanding Machine Learning: From Theory to Algorithms*, (Cambridge University Press, Cambridge, 2014).

PAC Learning

Definition.
A learning function

$$\mathcal{A}\colon \bigcup_{m\in\mathbb{N}} \mathcal{Z}^m \to \mathcal{H}$$

for $\mathcal{H}$ is said to be probably approximately correct (PAC) (with respect to $\mathcal{D}$) if it satisfies the following condition:

$$\forall \varepsilon, \delta \in (0,1)\ \exists m_0 \in \mathbb{N}\ \forall m \geq m_0\ \forall \mathbb{D} \in \mathcal{D}:$$
$$\mathbb{D}^m(\{\mathbf{z} \in \mathcal{Z}^m \mid \mathrm{er}_{\mathbb{D}}(\mathcal{A}(\mathbf{z})) - \mathrm{opt}_{\mathbb{D}}(\mathcal{H}) \leq \varepsilon\}) \geq 1 - \delta.$$

The hypothesis space $\mathcal{H}$ is said to be PAC learnable if there exists a learning function for $\mathcal{H}$ that is PAC.

L. G. Valiant, 'A Theory of the Learnable', *Comm. ACM* **27** (1984) 1134–1142.

PAC Learning – refined

Definition.

A learning function

$$\mathcal{A}\colon \bigcup_{m\in\mathbb{N}} \mathcal{Z}^m \to \mathcal{H}$$

for $\mathcal{H}$ is said to be probably approximately correct (PAC) (with respect to $\mathcal{D}$) if it satisfies the following condition:

$$\forall \varepsilon, \delta \in (0,1)\ \exists m_0 \in \mathbb{N}\ \forall m \geq m_0\ \forall \mathbb{D} \in \mathcal{D}\ \exists C \in \boldsymbol{\Sigma}_{\mathcal{Z}}^m :$$
$$C \subseteq \{\mathbf{z} \in \mathcal{Z}^m \mid \mathrm{er}_{\mathbb{D}}(\mathcal{A}(\mathbf{z})) - \mathrm{opt}_{\mathbb{D}}(\mathcal{H}) \leq \varepsilon\}$$
$$\text{and } \mathbb{D}^m(C) \geq 1 - \delta.$$

The hypothesis space $\mathcal{H}$ is said to be PAC learnable if there exists a learning function for $\mathcal{H}$ that is PAC.

Fundamental Theorem of Statistical Learning: *Agnostic* Version

Based on Blumer, Ehrenfeucht, Haussler and Warmuth 1989, we could prove the following:

Theorem.

Let $\mathcal{D}$ contain all discrete uniform distributions and let $\mathcal{H}$ be well-behaved with respect to $\mathcal{D}$. Then $\mathcal{H}$ is PAC learnable with respect to $\mathcal{D}$ if and only if $\mathcal{H}$ has finite VC dimension.

L. S. Krapp and L. Wirth, 'Measurability in the Fundamental Theorem of Statistical Learning', Preprint, 2024, arXiv:2410.10243.

Discrete Uniform Distributions

Definition

A discrete uniform distribution on a measurable space (Ω, Σ) with $\mathcal{P}_{\mathrm{fin}}(\Omega) \subseteq \Sigma$ is a probability measure $\mathbb{P}\colon \Sigma \to [0, 1]$ of the form

$$\mathbb{P} = \sum_{j=1}^{\ell} \frac{1}{\ell} \delta_{\omega_j},$$

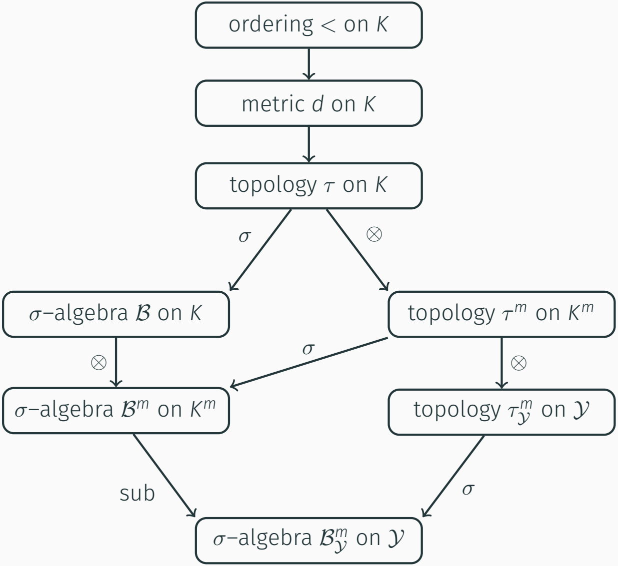
where $\ell \in \mathbb{N}$ and $\omega_1, \ldots, \omega_\ell \in \Omega$.

Measurability in Subfields of $\mathbb{R}$ – Borel σ–Algebras

Setting:

$K \subseteq \mathbb{R}$ subfield

$\mathcal{Y} \subseteq K^m$



Measurability in Subfields of $\mathbb{R}$ – Identity

Setting:

$K \subseteq \mathbb{R}$ subfield

$\mathcal{Y} \subseteq K^m$

$\mathcal{X} \subseteq K^n$

$\mathcal{Z} = \mathcal{X} \times \{0,1\} \subseteq K^{n+1}$