Generalised Entropies and Asymptotic Complexities of Languages

Revised and Updated

Yuri Kalnishkan, Michael V. Vyugin, and Vladimir Vovk

Department of Computer Science
and Computer Learning Research Centre
Royal Holloway, University of London

2012

Papers

- conference version:
  Y. Kalnishkan, V. Vovk and M. V. Vyugin. Generalised Entropy and Asymptotic Complexities of Languages. In Learning Theory, 20th Annual Conference on Learning Theory, *COLT 2007*, volume 4539 of Lecture Notes in Computer Science, pages 293-307, Springer 2007.
  — limited to two games
  — inaccuracies in the main result
- full version accepted for publication in *Information and Computation*

Talk Outline

Protocol

- we try to predict elements of a sequence $\omega_1, \omega_2, \omega_3, \ldots \in \Omega$
- we output predictions $\gamma_1, \gamma_2, \gamma_3, \ldots \in \Gamma$
- protocol:
  FOR $t = 1, 2, \ldots$
  (1) $\mathfrak{A}$ chooses a prediction $\gamma_t \in \Gamma$
  (2) $\mathfrak{A}$ observes the actual outcome $\omega_t \in \Omega$
  END FOR
- the quality of predictions is measured by a loss function $\lambda(\omega, \gamma)$
  — loss over T trials sums up to the cumulative loss

$$\mathrm{Loss}_{\mathfrak{A}}(\omega_1, \omega_2, \ldots, \omega_T) = \sum_{i=1}^{T} \lambda(\omega_i, \gamma_i)$$

Formalisation

- a *game* $\mathfrak{G}$ is a triple $\langle \Omega, \Gamma, \lambda \rangle$
  — Ω is the *outcome space*
  — Γ is the *prediction space*
  — $\lambda : \Omega \times \Gamma \to [0, +\infty]$ is the *loss function*
- in this talk
  — $\Omega = \{\omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(M-1)}\}$ is finite
  — Γ is compact
  — λ is continuous

Prediction Strategy

- $\mathfrak{A} : \Omega^* \to \Gamma$ maps finite sequences of previous outcomes to predictions
- we can consider various classes of strategies, e.g., computable and polynomial-time computable
  — but the ultimate goal is to study predictability

Loss as Complexity

- the loss of a strategy $\mathfrak{A}$ on a sequence $\boldsymbol{x} = (\omega_1, \omega_2, \dots, \omega_n)$ is

$$\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda(\omega_i, \mathfrak{A}(\omega_1, \omega_2, \dots, \omega_{i-1}))$$

- this can be thought of as complexity of $\boldsymbol{x}$ w.r.t. $\mathfrak{A}$
- can we define complexity irrespective of $\mathfrak{A}$?
  — if we take 'optimal' $\mathfrak{A}$, we can consider its loss as 'intrinsic' complexity of $\boldsymbol{x}$

Difficulties

- for every fixed sequence there is a strategy that knows it already
  — unless we take computability into account...
- every strategy is beaten by some other strategy on some sequences

Predictive Complexity

- a solution: *predictive complexity* [Vovk and Watkins, 1998]
  — a class of semi-computable semi-strategies is considered
  — it usually has an optimal element
  — we can define predictive complexity of a sequence up to a constant
- the theory of predictive complexity is very similar to Kolmogorov complexity
- existence for various classes of losses is partly an open problem

Asymptotic Complexity

- let us consider complexity of *languages* (= sets of strings) instead of individual sequences
- let us consider loss per element
- let us consider limits
- we get something like

$$\mathrm{AC}(L) = \inf_{\mathfrak{A}} \lim_{n \to +\infty} \max_{\boldsymbol{x} \in L \cap \Omega^n} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x})}{n}$$

Questions

- so

$$\mathrm{AC}(L) = \inf_{\mathfrak{A}} \lim_{n\to+\infty} \max_{\boldsymbol{x}\in L\cap\Omega^n} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x})}{n}$$

- what if there are no $\boldsymbol{x}$s of length n?
  — skip that n
- what if there are no $\boldsymbol{x}$s of length n from some length on?
  — no complexity for finite languages
- what if the limit does not exist?
  — let us consider upper and lower limits instead
- if the sequence is infinite, we can first take the limit $\lim_{n\to+\infty}$ along the sequence and then $\sup_{\boldsymbol{x}\in L}$
  — two more variations of complexity

Finite Sequences

- let $L \subseteq \Omega^*$ (L is a set of finite sequences)
  — let L be infinite
- *upper* (uniform) complexity:

$$\overline{\mathrm{AC}}(L) = \inf_{\mathfrak{A}} \limsup_{n\to+\infty} \max_{\boldsymbol{x}\in L\cap\Omega^n} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x})}{n}$$

- *lower* (uniform) complexity:

$$\underline{\mathrm{AC}}(L) = \inf_{\mathfrak{A}} \liminf_{n\to+\infty} \max_{\boldsymbol{x}\in L\cap\Omega^n} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x})}{n}$$

- in the former definition we assume $\max\varnothing = 0$ and in the later $\max\varnothing = +\infty$

Infinite Sequences

- let $L \subseteq \Omega^\infty$ (L is a set of infinite sequences)
- we can consider the set of all finite prefixes of all sequences from L; it has upper and lower complexities; let us call them *upper uniform* complexity $\overline{\mathrm{AC}}(L)$ and *lower uniform* complexity $\underline{\mathrm{AC}}(L)$
- *upper non-uniform* complexity:

$$\overline{\overline{\mathrm{AC}}}(L) = \inf_{\mathfrak{A}} \sup_{\boldsymbol{x}\in L} \limsup_{n\to+\infty} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}|_n)}{n}$$

- *lower non-uniform* complexity:

$$\underline{\underline{\mathrm{AC}}}(L) = \inf_{\mathfrak{A}} \sup_{\boldsymbol{x}\in L} \liminf_{n\to+\infty} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}|_n)}{n}$$

Problem

- suppose we have games $\mathfrak{G}_1, \mathfrak{G}_2, \ldots, \mathfrak{G}_K$ with same set of possible outcomes Ω
- what are the relations among $\mathrm{AC}_1, \mathrm{AC}_2, \ldots, \mathrm{AC}_K$?
- we shall describe the set $\{(\mathrm{AC}_1(L), \mathrm{AC}_2(L), \ldots, \mathrm{AC}_K(L))\}$ on $\mathbb{R}^K$
  — here L ranges over all non-trivial languages

Weak Mixability

- a game is weakly mixable if for any two strategies $\mathfrak{A}_1$ and $\mathfrak{A}_2$ there is a strategy $\mathfrak{A}$ such that

$$\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}) \leq \min\left(\mathrm{Loss}_{\mathfrak{A}_1}(\boldsymbol{x}), \mathrm{Loss}_{\mathfrak{A}_2}(\boldsymbol{x})\right) + o(|\boldsymbol{x}|)$$

Convexity

- let us take the set
$P = \{(\lambda(\omega^{(0)}, \gamma), \lambda(\omega^{(1)}, \gamma), \ldots, \lambda(\omega^{(M-1)}, \gamma)) \mid \gamma \in \Gamma\} \subseteq \mathbb{R}^M$
— images of points from Γ in $\mathbb{R}^M$
- a point $(s_0, s_1, \ldots, s_{M-1}) \in \mathbb{R}^M$ is a *superprediction* if there is $p = (p_0, p_1, \ldots, p_{M-1})$ such that

$$\begin{aligned} p_0 &\leq s_0 \\ p_1 &\leq s_1 \\ &\cdots \\ p_{M-1} &\leq s_{M-1} \end{aligned}$$

- superpredictions are located ‘above and to the right’ from points of P
- weak mixability is equivalent to convexity of S

Entropy

- let p be a probability distribution on Ω
— $p = (p_0, p_1, \ldots, p_{M-1})$, where $\sum p_i = 1$
- *generalised entropy*

$$H(p) = \min_{\gamma \in \Gamma} \mathbf{E}_p \lambda(\omega, \gamma) = \min_{\gamma \in \Gamma} \sum_{i=0}^{M-1} p_i \lambda(\omega^{(i)}, \gamma)$$

— suppose we know that the next outcome is distributed according to P
— we will be looking for $\gamma \in \Gamma$ to minimise the expected loss
— the minimum of the expected loss is the entropy $H(p)$
- discussed in [Grünwald and Dawid, 2004]

Entropy Hull

- suppose we have games $\mathfrak{G}_1, \mathfrak{G}_2, \ldots, \mathfrak{G}_K$ (with the same Ω)
  — they specify entropies $H_1, H_2, \ldots, H_K$
- consider the set
  $\{(H_1(p), H_2(p), \ldots, H_K(p)) \mid p \text{ is a distribution}\}$
- $\mathfrak{G}_1/\mathfrak{G}_2/\ldots/\mathfrak{G}_K$*-entropy hull* is its convex hull
- this is nearly the solution to our problem...

Lattices

- a set $M \subseteq \mathbb{R}^K$ is a *sublattice* of $\mathbb{R}^K$ if for every two points $x, y \in M$

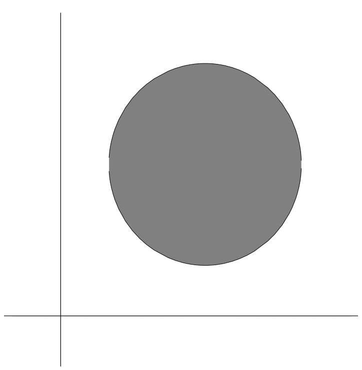
$$\max(x, y) \in M$$
$$\min(x, y) \in M$$

- a set $M \subseteq \mathbb{R}^K$ is a *upper subsemisublattice* of $\mathbb{R}^K$ if for every two points $x, y \in M$

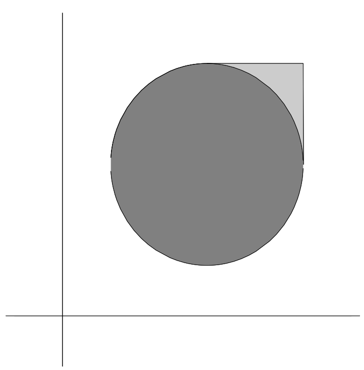
$$\max(x, y) \in M$$

- a sublattice closure is the smallest sublattice of $\mathbb{R}^K$ containing M
  — upper subsemilattice closure is defined similarly

Example



M is not a lattice



the upper subsemilattice closure of M

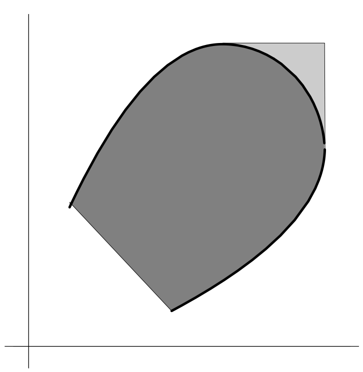
Main Theorem (1)

- let $\mathfrak{G}_1, \mathfrak{G}_2, \ldots, \mathfrak{G}_K$ be weakly mixable with the same set of possible outcomes
- then the set of all pairs $(\mathrm{AC}_1(L), \mathrm{AC}_2(L), \ldots, \mathrm{AC}_K(L))$, where
  — AC is one of the complexities $\overline{\mathrm{AC}}$ or $\overline{\overline{\mathrm{AC}}}$
  — L ranges over all non-empty sets of infinite sequences or all infinite sets of finite sequences accordingly
- coincides with the upper subsemilattice closure of the $\mathfrak{G}_1/\mathfrak{G}_2/\ldots/\mathfrak{G}_K$-entropy hull

Main Theorem (2)

- the set of all pairs $(\mathrm{AC}_1(L), \mathrm{AC}_2(L), \ldots, \mathrm{AC}_K(L))$, where
  — AC is one of the complexities $\underline{\mathrm{AC}}$ or $\underline{\underline{\mathrm{AC}}}$
  — L ranges over all non-empty sets of infinite sequences or all infinite sets of finite sequences accordingly
- coincides with the sublattice closure of the $\mathfrak{G}_1/\mathfrak{G}_2/\ldots/\mathfrak{G}_K$-entropy hull

Building the Set



entropy curve $\to$ entropy hull $\to$ upper subsemilattice closure

1. On-line Prediction
2. Complexities
3. Preliminaries
4. Main Result
5. Proof Sketch

Recalibration

- let $\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_K$ be strategies for weakly mixable games $\mathfrak{G}_1, \mathfrak{G}_2, \ldots, \mathfrak{G}_K$
- then for every weakly mixable $\mathfrak{G}$ and $\varepsilon > 0$ there is a strategy $\mathfrak{S}$ and a function $f(n) = o(n)$ as $n \to \infty$ such that for every finite string $\boldsymbol{x} \in \Omega^*$ there are
  — distributions $p_1, p_2, \ldots, P_N$
  — distribution $q = (q_1, q_2, \ldots, q_N)$ such that

$$\sum_{i=1}^{N} q_i H_k(p_i) \leq \frac{\mathrm{Loss}_{\mathfrak{A}_k}^{\mathfrak{G}_k}(\boldsymbol{x})}{|\boldsymbol{x}|} + \varepsilon$$

$$\mathrm{Loss}_{\mathfrak{S}}^{\mathfrak{G}}(\boldsymbol{x}) \leq |\boldsymbol{x}| \left(\sum_{i=1}^{N} q_i H(p_i) + \varepsilon \right) + f(|\boldsymbol{x}|)$$