Generalized entropies and asymptotic complexities of languages

Yuri Kalnishkan, Vladimir Vovk, Michael Vyugin

Department of Computer Science
Royal Holloway, University of London
Egham, Surrey, England

{yura,vovk,misha}@cs.rhul.ac.uk

San Diego, June 14, 2007

Games (of prediction)

A game $\mathfrak{G}$ is a triple $\langle \Omega, \Gamma, \lambda \rangle$, where Ω is an outcome space, Γ is a prediction space, and $\lambda : \Omega \times \Gamma \to [0, \infty]$ is a loss function.

This talk (and paper): comparing different games (are they close or are they very different?).

Assumptions

- $\Omega = \{\omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(M-1)}\}$ is finite. The binary case: $M = 2$; Ω is identified with $\mathbb{B} = \{0, 1\}$.
- Γ is a compact topological space.
- $\lambda(\omega, \gamma)$ is continuous in γ.
- The set $\{\gamma \in \Gamma \mid \lambda(\omega, \gamma) < \infty, \forall \omega\}$ is dense in Γ.

Important examples

Three binary games with $\Omega = \mathbb{B}$ and $\Gamma = [0, 1]$:

- the square-loss game with $\lambda(\omega, \gamma) = (\omega - \gamma)^2$;
- the absolute-loss game with $\lambda(\omega, \gamma) = |\omega - \gamma|$;
- the logarithmic game with

$$\lambda(\omega, \gamma) = \begin{cases} -\log(1-\gamma) & \text{if} \quad \omega = 0 \\ -\log\gamma & \text{if} \quad \omega = 1. \end{cases}$$

Strategies and losses

A prediction strategy: $\mathfrak{A} : \Omega^* \to \Gamma$. On a finite sequence $\boldsymbol{x} = x_1 x_2 \dots x_n \in \Omega^n$ it suffers loss

$$\mathrm{Loss}^{\mathfrak{G}}_{\mathfrak{A}}(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda(x_i, \mathfrak{A}(x_1 x_2 \dots x_{i-1})).$$

An M-tuple $(s_0, s_1, \dots, s_{M-1}) \in [0,\infty]^M$ is a superprediction w.r. to $\mathfrak{G}$ if there is a prediction $\gamma \in \Gamma$ such that $\lambda(\omega^{(i)}, \gamma) \le s_i$ for all $i = 0, 1, \dots, M-1$. The game $\mathfrak{G}$ is convex if the finite part of its set of superpredictions ($S \cap \mathbb{R}^M$, where S is the set of superpredictions) is convex.

Generalized Entropies

Fix a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$. Let $\mathbb{P}(\Omega)$ be the set of probability distributions on Ω. Since Ω is finite, we can identify $\mathbb{P}(\Omega)$ with the standard $(M-1)$-simplex $\mathbb{P}_M = \{(p_0, p_1, \ldots, p_{M-1}) \in [0,1]^M \mid \sum_{i=0}^{M-1} p_i = 1\}$.

Generalized entropy for $p^* = (p_0, p_1, \ldots, p_{M-1}) \in \mathbb{P}(\Omega)$:

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

with the convention $0 \times \infty = 0$.

Binary case

In the binary case: all information about $p^* \in \mathbb{P}(\mathbb{B})$ is contained in p, the probability of 1, and so $\mathbb{P}(\mathbb{B})$ can be identified with $[0, 1]$;

$$H(p) = \min_{\gamma \in \Gamma} \left((1-p)\lambda(0,\gamma) + p\lambda(1,\gamma)\right).$$

- for the square-loss game, $H(p) = p(1-p)$;
- for the absolute-loss game, $H(p) = \min(p, 1-p)$;
- for the logarithmic game, $H(p) = -p \log p - (1-p) \log(1-p)$.

Asymptotic complexities for finitary languages

Fix a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$. Let $L \subseteq \Omega^*$ (finitary language).

The upper and lower asymptotic complexity of L w.r.t. to the game $\mathfrak{G}$:

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

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

with the convention $\max\varnothing = 0$ in the first and $\max\varnothing = \infty$ in the second.

Asymptotic complexities for infinitary languages

We consider two natural ways to define complexities of languages $L \subseteq \Omega^\infty$:

- Identify L with the set of all finite prefixes of all its sequences. We use the same notation $\overline{\mathsf{AC}}(L)$ and $\underline{\mathsf{AC}}(L)$ and refer to these complexities as uniform.
- Non-uniform complexities:

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

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

Other terminology

Lutz and Fortnow's dimension = the lower non-uniform complexity w.r. to the multidimensional version of the logarithmic game; their predictability $= 1 - \underline{\underline{\mathrm{AC}}}$, where $\underline{\underline{\mathrm{AC}}}$ is the lower non-uniform complexity w.r. to the multidimensional version of the absolute-loss game.

Differences between complexities

Simple examples show that the introduced complexities are different:

- Upper and lower complexities differ.
- Uniform complexities differ from non-uniform.

Main Result

Consider games $\mathfrak{G}_1$ and $\mathfrak{G}_2$ with the same finite Ω and entropies H_1 and H_2. The $\mathfrak{G}_1/\mathfrak{G}_2$-entropy set is $\{(H_1(p), H_2(p)) \mid p \in \mathbb{P}(\Omega)\}$. Its convex hull: the $\mathfrak{G}_1/\mathfrak{G}_2$-entropy hull.

A closed convex $\mathcal{S} \subseteq \mathbb{R}^2$ is a rocket if

$$(x_1, y_1), (x_2, y_2) \in \mathcal{S} \Longrightarrow (\max(x_1, x_2), \max(y_1, y_2)) \in \mathcal{S}.$$

The rocket closure of $\mathcal{H} \subseteq \mathbb{R}^2$ is the smallest rocket containing $\mathcal{H}$.

Theorem If games $\mathfrak{G}_1$ and $\mathfrak{G}_2$ have the same finite outcome space Ω and are convex, then the rocket closure of the $\mathfrak{G}_1/\mathfrak{G}_2$-entropy hull coincides with the following sets:

- $\{(\overline{\mathsf{AC}}_1(L), \overline{\mathsf{AC}}_2(L)) \mid L \subseteq \Omega^* \text{ and } L \text{ is infinite}\}$;
- $\{(\underline{\mathsf{AC}}_1(L), \underline{\mathsf{AC}}_2(L)) \mid L \subseteq \Omega^* \text{ and } L \text{ is infinite}\}$;
- $\{(\overline{\mathsf{AC}}_1(L), \overline{\mathsf{AC}}_2(L)) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing\}$;
- $\{(\underline{\mathsf{AC}}_1(L), \underline{\mathsf{AC}}_2(L)) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing\}$;
- $\{(\overline{\overline{\mathsf{AC}}}_1(L), \overline{\overline{\mathsf{AC}}}_2(L)) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing\}$;
- $\{(\underline{\underline{\mathsf{AC}}}_1(L), \underline{\underline{\mathsf{AC}}}_2(L)) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing\}$.

Remarks

The last item on the list covers Lutz and Fortnow's COLT'2002 result.

Simple examples show that:

- The requirement of convexity cannot be omitted.
- The statement of the theorem does not apply to pairs $(\overline{\mathsf{AC}}_1(L), \underline{\mathsf{AC}_2}(L))$ or pairs $(\overline{\overline{\mathsf{AC}}}_1(L), \underline{\underline{\mathsf{AC}}}_2(L))$.

Alternative formalization (informal)

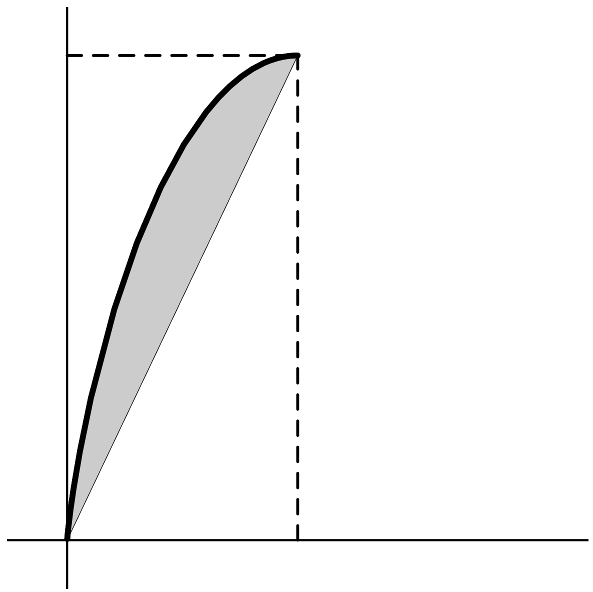
Given a game $\mathfrak{G}$ and $\boldsymbol{x} \in \Omega^*$, the predictive complexity $K^{\mathfrak{G}}(\boldsymbol{x})$ is the smallest loss suffered by "semicomputable prediction strategies"; defined to within $O(1)$ for "perfectly mixable" games (such as SQ and LOG) and to within $O\left(\sqrt{|\boldsymbol{x}|}\right)$ for bounded and convex games (such as SQ and ABS).

For each n, plot the set of $\left(K^{\mathfrak{G}_1}(\boldsymbol{x}), K^{\mathfrak{G}_2}(\boldsymbol{x})\right)$, $\boldsymbol{x} \in \Omega^n$. Go to the limit as $n \to \infty$. The resulting set will coincide with the $\mathfrak{G}_1/\mathfrak{G}_2$-entropy hull.

Idea of the proof

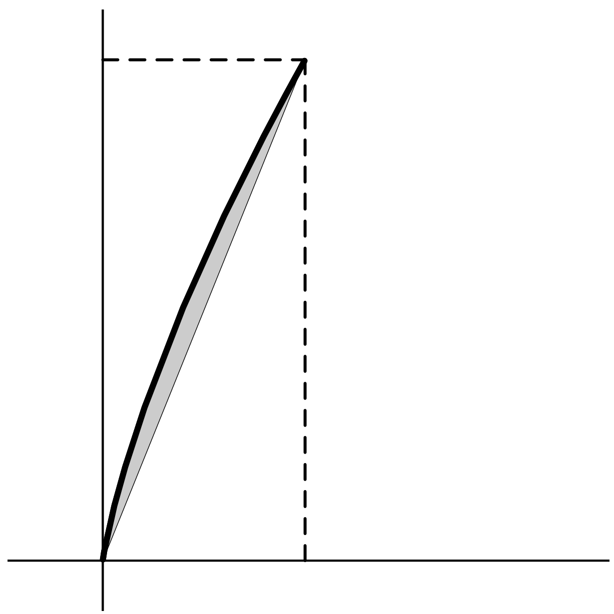
- recalibration (permutation of a grid in Γ^2)
- random generation

The ABS/LOG convex hull



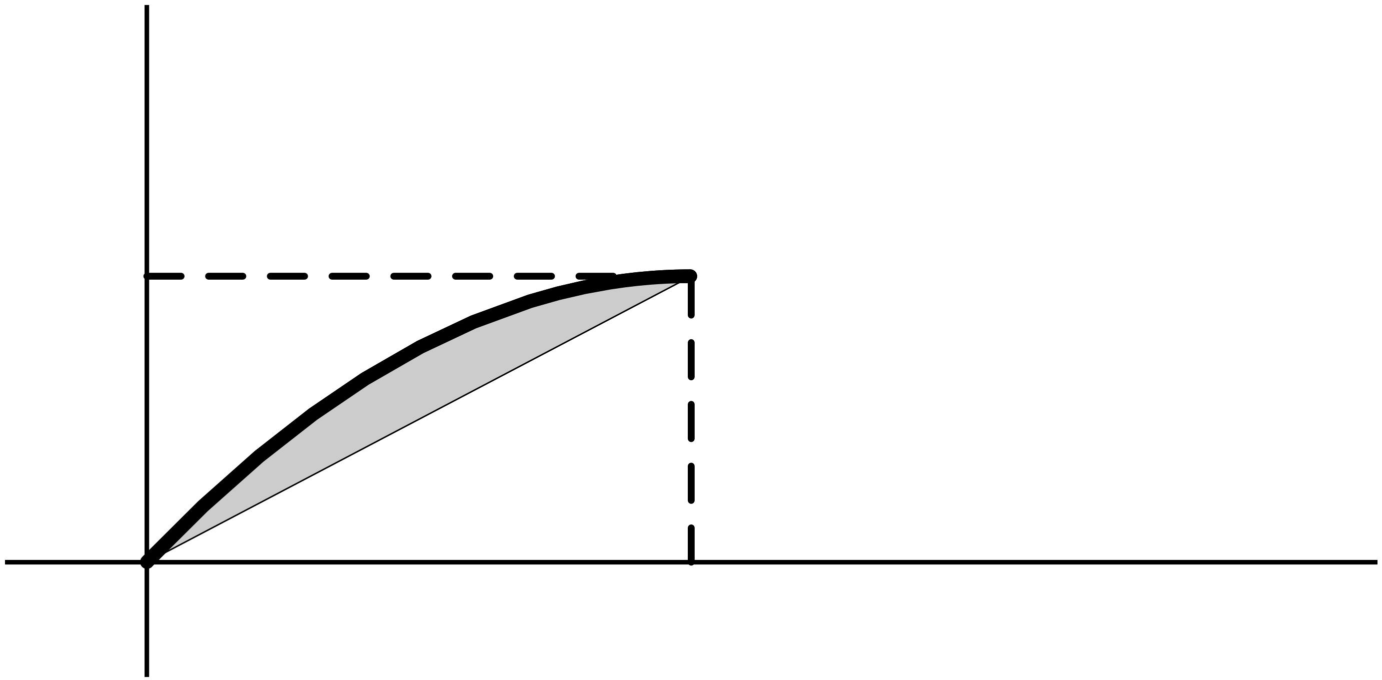
The difference is significant.

The SQ/LOG convex hull



Look more similar.

The ABS/SQ convex hull



Again significant difference.