Predictive complexity, information and randomness

Vladimir Vovk

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

vovk@cs.rhul.ac.uk

Authors of the results: Yuri Kalnishkan, Michael Vyugin, VV.

This talk (and related papers): http://www.vovk.net.

A possible view of Kolmogorov complexity: $K(z)$ is the optimal loss suffered when predicting the elements of $z = \omega_1 \dots \omega_n$ on-line. This talk:

- the on-line prediction problem; predictability classes
- universal prediction algorithms; predictive complexity and Kolmogorov complexity ($KM(z) = -\log M(z)$) as its special case; criterion of existence of predictive complexity in the binary case
- predictive information (definition)
- predictive randomness; unpredictability and its special case, incompressibility

"Prediction": coding (Kolmogorov complexity), investing in a stock market, prediction.

Games of prediction

FOR $i = 1, 2, \ldots$ Forecaster

(1) chooses prediction $\gamma_i \in \Gamma$

(2) observes actual outcome $\omega_i \in \Omega$

(3) suffers loss $\lambda(\gamma_i, \omega_i)$

END FOR

Main example

Log-loss game LOG:

$$\Omega = \{0, 1\},\ \Gamma = [0, 1],$$

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

Game of coding (or probability forecasting).

Three more examples

Square-loss game SQ:

$$\Omega = \{0,1\},\ \Gamma = [0,1],$$

$$\lambda(\gamma,\omega) = (\gamma-\omega)^2;$$

binary-loss game BIN:

$$\Omega = \{0,1\},\ \Gamma = \{0,1\},$$

$$\lambda(\gamma,\omega) = |\gamma-\omega|;$$

absolute-loss game ABS:

$$\Omega = \{0,1\},\ \Gamma = [0,1],$$

$$\lambda(\gamma,\omega) = |\gamma-\omega|$$

(the expected loss in the binary-loss game).

If S is a prediction strategy, $\mathrm{Loss}_S^\lambda(z)$ is the loss incurred by Forecaster who follows S after Nature chooses $z = (\omega_1, \ldots, \omega_n)$. Let $\Omega := \{0, 1\}$.

Language: infinite $L \subseteq \{0, 1\}^*$; $L_n := L \cap \{0, 1\}^n$.

If $\mathcal{F}$ is a class of functions $f : \mathbb{N} \to \mathbb{N}$, $\mathbf{C}$ is a computational complexity class and λ is a loss function,

$$\mathbf{PRED}^\lambda(\mathcal{F}, \mathbf{C})$$

is the class of languages L for which $\exists\ S \in \mathbf{C}$ and $f \in \mathcal{F}$ such that

$$\forall n \forall z \in L_n : \mathrm{Loss}_S^\lambda(z) \leq f(n).$$

Examples:

$$\mathbf{PRED}^{\mathsf{LOG}}(0.01n + o(n), \mathbf{ALL}).$$

$$\mathbf{PRED}^{\mathsf{SQ}}(o(n), \mathbf{COMP}),$$

$$\mathbf{PRED}^{\mathsf{ABS}}(o(n), \mathbf{P}),$$

$$\mathbf{PRED}^{\mathsf{BIN}}(O(\log n), \mathbf{L}),$$

Big problem: Study the structure of the predictability classes $\mathbf{PRED}^{\lambda}(\mathcal{F}, \mathbf{C})$ w.r. to $\subseteq$.

Simple special case: weakly mixable games and $\mathbf{C} = \mathbf{ALL}$. (Fortnow and Lutz: LOG vs. ABS, $\mathbf{C} = \mathbf{P}$, languages $\mapsto$ infinite sequences.)

λ is weakly mixable if

$$\exists f(n) = o(n) \forall S_1, S_2 \exists S \forall n \forall z \in \{0,1\}^n :$$

$$\mathsf{Loss}_S^\lambda(z) \le \min\left(\mathsf{Loss}_{S_1}^\lambda(z), \mathsf{Loss}_{S_2}^\lambda(z)\right) + f(n).$$

Examples: LOG, SQ, ABS; mixing can be done using the "Aggregating Algorithm".

Set

$$\mathbf{SP}^\lambda(c) := \mathbf{PRED}^\lambda(cn + o(n), \mathbf{ALL}).$$

λ-entropy:

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

[determines λ up to parameterization by the Fenchel-Moreau theorem]

Examples: log-entropy (=entropy)

$$H(p) = -p \log p - (1-p) \log(1-p),$$

square-loss entropy

$$H^{\mathrm{SQ}}(p) = p(1-p),$$

absolute-loss (=binary-loss) entropy

$$H^{\mathrm{ABS}}(p) = \min(p, 1-p).$$

λ_1/λ_2-entropy curve:

$$\left\{ \left(H^{\lambda_1}(p), H^{\lambda_2}(p) \right) : p \in [0,1] \right\}.$$

E.g.: the ABS/LOG-entropy curve is the left-hand half of H.

λ_1/λ_2-entropy hull := the convex hull of the λ_1/λ_2-entropy curve.

Theorem 1. Suppose λ_1 and λ_2 are weakly mixable. Then

$$\mathbf{SP}^{\lambda_1}(c_1) \subseteq \mathbf{SP}^{\lambda_2}(c_2)$$

iff, for all (x, y) in the λ_1/λ_2-entropy hull,

$$x \leq c_1 \Longrightarrow y \leq c_2.$$

Slightly stronger (and more intuitive) statement: setting

$$\mathbf{AC}^{\lambda}(L) := \text{inf}\left\{c : L \in \mathbf{SP}^{\lambda}(c)\right\},$$

the set

$$\left\{\left(\mathbf{AC}^{\lambda_1}(L), \mathbf{AC}^{\lambda_2}(L)\right) : L \text{ is a language}\right\}$$

coincides with the closure of the λ_1/λ_2-entropy hull, where $A \subseteq \mathbb{R}^2$ being closed means

$$(x_1, y_1), (x_2, y_2) \in A \implies (x_1 \vee x_2, y_1 \vee y_2) \in A.$$

Standard regularity conditions:

- Γ is a compact topological space
- $\lambda(\gamma, \omega)$ is continuous in γ
- $\sup_\omega \lambda(\gamma, \omega) < \infty$ for some γ
- $\sup_\omega \lambda(\gamma, \omega) > 0$ for all γ

Predictive complexity

From now on: λ is assumed computable.

We are interested in $\text{Loss}_S^\lambda(z)$ (loss process of S) for the best computable S. Such S does not exist!

$\mathcal{L} : \Omega^* \to \mathbb{R}$ is a measure of predictive complexity if:

- $\mathcal{L}$ is a superloss process: $\mathcal{L}(\Box) = 0$ and
$$\forall z \, \exists \gamma \, \forall \omega : \mathcal{L}(z * \omega) \geq \mathcal{L}(z) + \lambda(\gamma, \omega);$$
- $\mathcal{L}$ is semicomputable from above: there exists a computable sequence of computable functions f_k such that $\mathcal{L} = \inf_k f_k$.

A smallest, to within an additive constant, measure $\mathcal{K}$ of predictive complexity is universal;

$$\forall \mathcal{L}\, \exists C\, \forall z : \mathcal{K}(z) \leq \mathcal{L}(z) + C.$$

If it exists: predictive complexity.

Theorem 2. Predictive complexity exists if and only if the prediction game is mixable.

A game is mixable if

$$\exists C \forall S_1, S_2 \exists S \forall z \in \{0,1\}^* :$$

$$\mathrm{Loss}^{\lambda}_{S}(z) \le \min\left(\mathrm{Loss}^{\lambda}_{S_1}(z), \mathrm{Loss}^{\lambda}_{S_2}(z)\right) + C.$$

There is an easier-to-check definition.

Canonical representation (CR) of a game:

$$\{(\lambda(\gamma,0), \lambda(\gamma,1)) : \gamma \in \Gamma\} \subseteq [0,\infty]^2$$

(the game to within parameterization). For an exponential learning rate $\beta = e^{-\eta}$, the exponential representation:

$$\left\{\left(\beta^{\lambda(\gamma,0)}, \beta^{\lambda(\gamma,1)}\right) : \gamma \in \Gamma\right\} \subseteq [0,1]^2.$$

A game is mixable iff for some $\beta \in (0,1)$ the exponential representation is concave.

Bounded λ with convex CR: weakly mixable (with $f(n) = O(\sqrt{n})$).

Proof of Theorem 2:

- Predictive complexity exists for mixable games: simple (AA: you can mix in the exponential picture).
- If predictive complexity exists, the game is mixable: more complicated.

This is what becomes with

$$\exists \gamma \, \forall \omega : \mathcal{L}(z * \omega) \geq \mathcal{L}(z) + \lambda(\gamma, \omega)$$

in the log-loss case:

$$\exists \gamma : \mathcal{L}(z{*}0) \geq \mathcal{L}(z) - \log(1-\gamma) \,\&\, \mathcal{L}(z{*}1) \geq \mathcal{L}(z) - \log \gamma;$$

$$\exists \gamma : 2^{-\mathcal{L}(z*0)} \leq (1-\gamma) 2^{-\mathcal{L}(z)} \,\&\, 2^{-\mathcal{L}(z*1)} \leq \gamma 2^{-\mathcal{L}(z)};$$

$$2^{-\mathcal{L}(z*0)} + 2^{-\mathcal{L}(z*1)} \leq 2^{-\mathcal{L}(z)}.$$

Therefore:

$$\mathcal{K}(z) =^{+} KM(z) := -\log M(z);$$

close to $K(z)$ and $C(z)$.

The game is obviously mixable.

Mixable games: log-loss, square-loss.

Binary-loss and absolute-loss game: not mixable. Therefore, predictive complexity does not exist for them.

Now assume $\lambda \geq 0$.

In the absolute-loss game,

$$\exists \mathcal{K} \, \forall \mathcal{L} \, \exists C \, \forall z : \mathcal{K}(z) \le \mathcal{L}(z) + C\sqrt{n},$$

where $n = l(z)$ is the length of z.

In the binary-loss game,

$$\exists \mathcal{K} \, \forall \mathcal{L} \, \exists C \, \forall z : \mathcal{K}(z) \le C\mathcal{L}(z).$$

Hypothesis: In the bounded binary ($\Omega = \{0, 1\}$) case, predictive complexity exists with accuracy $\cdots + O(1)$ (mixable games), $\cdots + O(n^{1/2})$, or $O(\cdots)$; no other accuracy is possible.

Open problem: Find a general algorithm determining the accuracy with which predictive complexity exists.

Continuous games

Square-loss and absolute-loss: set $\Omega := [0, 1]$.

Natural extensions of the log-loss game:
Kullback-Leibler game and Cover's game.

Kullback–Leibler game:

$$\Omega = \Gamma = [0, 1],$$

$$\lambda(\gamma, \omega) = \omega \log \frac{\omega}{\gamma} + (1 - \omega) \log \frac{1 - \omega}{1 - \gamma}.$$

Cover's game with 2 stocks (0 and 1):

$$\Omega = [0, \infty)^2, \ \Gamma = [0, 1],$$

$$\lambda(\gamma, (\omega_0, \omega_1)) = -\log(\gamma\omega_1 + (1-\gamma)\omega_0).$$

The game is mixable and $\mathcal{K}^{\text{Cover}}$ contains $\mathcal{K}^{\text{LOG}}$, e.g.,

$$\mathcal{K}^{\text{LOG}}(001110) =$$

$$\mathcal{K}^{\text{Cover}}((1,0),(1,0),(0,1),(0,1),(0,1),(1,0)).$$

Long investment game.

All these continuous games are mixable.

Predictive information:

$$I(a : \omega_1 \dots \omega_n) :=$$
$$\mathcal{K}(\omega_1 \dots \omega_n) - \mathcal{K}(\omega_1 \dots \omega_n \mid a).$$

E.g.: how much message a increases the capital that can be made investing in a stock market represented by the sequence of stock prices $\omega_1 \dots \omega_n$?

Data-snooping curve:

$$S_z(\alpha) := \max_{a : l(a) \le \alpha} I(a : z).$$

E.g.: the maximal value of α bits of insider information.

Problem. Consider the binary case ($\Omega = \{0, 1\}$). Let f be the scaled version of S,

$$f_z(t) := \frac{S_z(nt)}{n}, \; t \in [0, 1].$$

What are the limiting points of sequences of functions f_{z_i}, $i = 1, 2, \ldots$, $l(z_i) = i$?

Closely related to the "Kolmogorov structure function". Partial results: Levin and Vereshchagin&Vitányi (log-loss case), Vladimir V'yugin.

Predictive randomness

Assume predictive complexity exists.

A finite sequence $z \in \Omega^n$ is C-random w.r. to a computable prediction strategy S (or S is C-efficient for z) if

$$\mathcal{K}(z) \geq \text{Loss}_S(z) - C.$$

For the log-loss game: $\approx$ Kolmogorov randomness.

Suppose $\Omega = \{0, 1\}$, CR is symmetric and $(B, B) \in$ CR. Trivial upper bound: $\mathcal{K}(z) \leq Bl(z)$.

Theorem 3. If CR is smooth near (B, B),

$$\sup_{n,m} \frac{\#\{z \in \{0,1\}^n : \mathcal{K}(z \mid m) \le Bn - m\}}{2^n \beta_0^m} \le 1$$

and

$$\inf_m \varliminf_{n\to\infty} \frac{\#\{z \in \{0,1\}^n : \mathcal{K}(z \mid m) \le Bn - m\}}{2^n \beta_0^m} > 0,$$

where β_0 is the smallest β such that the β-exponential representation lies below $x + y = 2\beta^B$.

For the log-loss game: incompressibility of Kolmogorov complexity.

Problem: Study unpredictability for arbitrary loss functions, probability distributions and predictive strategies.