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 $z = \omega_1 \ldots \omega_n$ 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, $\text{Loss}_S^\lambda(z)$ 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 $F$ is a class of functions $f : \mathbb{N} \rightarrow \mathbb{N}$, $C$ is a computational complexity class and $\lambda$ is a loss function,

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

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

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

\[ \text{PRED}^{\text{LOG}}(0.01n + o(n), \text{ALL}). \]

\[ \text{PRED}^{\text{SQ}}(o(n), \text{COMP}), \]

\[ \text{PRED}^{\text{ABS}}(o(n), P), \]

\[ \text{PRED}^{\text{BIN}}(O(\log n), L), \]

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

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

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

\[
\text{Loss}_S^\lambda(z) \leq \min (\text{Loss}_{S_1}^\lambda(z), \text{Loss}_{S_2}^\lambda(z)) + f(n).
\]

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

Set

\[
\text{SP}^\lambda(c) := \text{PRED}^\lambda(cn + o(n), \text{ALL}).
\]
**λ-entropy:**

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

[determines \( \lambda \) up to parameterization by the Fenchel-Moreau theorem]

**Examples:** log-entropy (\( \equiv \)entropy)

\[ H(p) = -p \log p - (1 - p) \log(1 - p), \]

square-loss entropy

\[ H^{\text{SQ}}(p) = p(1 - p), \]

absolute-loss (\( \equiv \)binary-loss) entropy

\[ H^{\text{ABS}}(p) = \min(p, 1 - p). \]
$\lambda_1/\lambda_2$-entropy curve:
$$\left\{(H^{\lambda_1}(p), H^{\lambda_2}(p)) : p \in [0, 1]\right\}.$$  
E.g.: the ABS/LOG-entropy curve is the left-hand half of $H$.

$\lambda_1/\lambda_2$-entropy hull := the convex hull of the $\lambda_1/\lambda_2$-entropy curve.

**Theorem 1.** Suppose $\lambda_1$ and $\lambda_2$ are weakly mixable. Then
$$\text{SP}^{\lambda_1}(c_1) \subseteq \text{SP}^{\lambda_2}(c_2)$$
iiff, for all $(x, y)$ in the $\lambda_1/\lambda_2$-entropy hull,
$$x \leq c_1 \implies y \leq c_2.$$
Slightly stronger (and more intuitive) statement: setting

\[ \text{AC}^\lambda(L) := \inf \{ c : L \in \text{SP}^\lambda(c) \}, \]

the set

\[ \{ (\text{AC}^{\lambda_1}(L), \text{AC}^{\lambda_2}(L)) : L \text{ is a language} \} \]

coincides with the closure of the \( \lambda_1/\lambda_2 \)-entropy hull, where \( A \subseteq \mathbb{R}^2 \) being closed means

\[(x_1, y_1), (x_2, y_2) \in A \implies (x_1 \lor x_2, y_1 \lor y_2) \in A.\]
Standard regularity conditions:

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

From now on: $\lambda$ is assumed computable.

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

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

- $L$ is a superloss process: $L(\Box) = 0$ and
  \[
  \forall z \exists \gamma \forall \omega : L(z \ast \omega) \geq L(z) + \lambda(\gamma, \omega);
  \]

- $L$ is semicomputable from above: there exists a computable sequence of computable functions $f_k$ such that $L = \inf_k f_k$. 

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

$$\exists C \forall S_1, S_2 \forall z \in \{0, 1\}^*: \quad \text{Loss}_S^\lambda(z) \leq \min \left( \text{Loss}_{S_1}^\lambda(z), \text{Loss}_{S_2}^\lambda(z) \right) + C.$$ 

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 $\lambda$ 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:

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

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,

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

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

In the binary-loss game,

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

**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:

\[ I(a : \omega_1 \ldots \omega_n) := \mathcal{K}(\omega_1 \ldots \omega_n) - \mathcal{K}(\omega_1 \ldots \omega_n | a). \]

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 \( \omega_1 \ldots \omega_n \)?

Data-snooping curve:

\[ S_z(\alpha) := \max_{a : l(a) \leq \alpha} I(a : z). \]

E.g.: the maximal value of \( \alpha \) bits of insider information.
Problem. Consider the binary case $(\Omega = \{0, 1\})$. Let $f$ be the scaled version of $S$,

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

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

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

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

Suppose $\Omega = \{0, 1\}$, CR is symmetric and $(B, B) \in \text{CR}$. Trivial upper bound: $K(z) \leq Bl(z)$. 
Theorem 3. If CR is smooth near \((B, B)\),

\[
\sup_{n,m} \frac{\# \{ z \in \{0, 1\}^n : \mathcal{K}(z \mid m) \leq Bn - m \}}{2^n \beta^m_0} \leq 1
\]

and

\[
\inf_m \lim_{n \to \infty} \frac{\# \{ z \in \{0, 1\}^n : \mathcal{K}(z \mid m) \leq Bn - m \}}{2^n \beta^m_0} > 0,
\]

where \(\beta_0\) is the smallest \(\beta\) such that the \(\beta\)-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.