NAFIT: New Algorithmic Forms of Information Theory

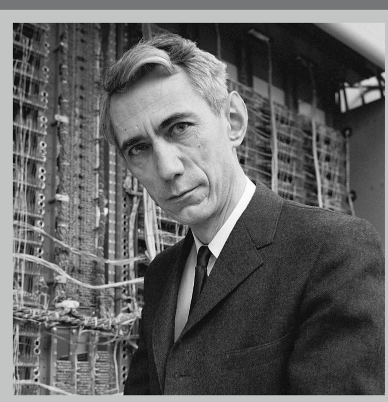
Gregory Lafitte and Alexander Shen

LIRMM – CNRS, Université de Montpellier II
Poncelet Lab. – Moscow Lomonosov University

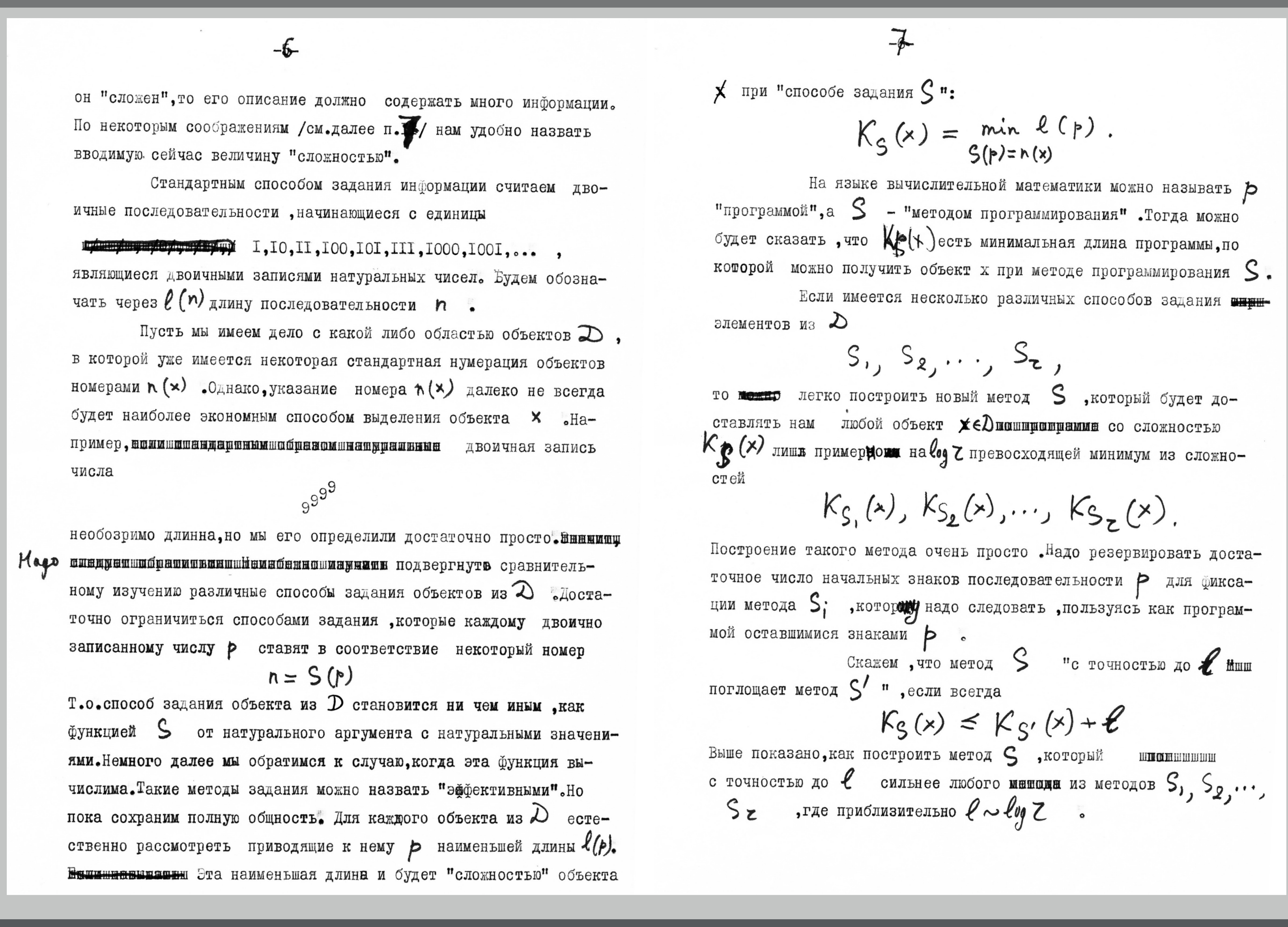
Measuring information

- **Classical information theory**: For a random variable that has $\mathbf{n}$ values with probabilities $\mathbf{p_1, \ldots, p_n}$, its *Shannon entropy* is defined as $\sum \mathbf{p_i} \log(1/\mathbf{p_i})$.
- **Algorithmic information theory**: The *Kolmogorov complexity* of a finite object $\mathbf{x}$ is the minimal length of a program that produces $\mathbf{x}$.

$$K_S(x) = \min_{S(p)=n(x)} \ell(p)$$



A page from Kolmogorov's autograph



-6-

он "сложен", то его описание должно содержать много информации. По некоторым соображениям /см. далее п. 7/ нам удобно назвать вводимую сейчас величину "сложностью".

Стандартным способом задания информации считаем двоичные последовательности, начинающиеся с единицы

I, IO, II, IOO, IOI, III, IOOO, IOOI, ... ,

являющиеся двоичными записями натуральных чисел. Будем обозначать через $\ell(n)$ длину последовательности n.

Пусть мы имеем дело с какой либо областью объектов $\mathcal{D}$, в которой уже имеется некоторая стандартная нумерация объектов номерами $n(x)$. Однако, указание номера $n(x)$ далеко не всегда будет наиболее экономным способом выделения объекта x. Например, двоичная запись числа

$$9^{9^{9^9}}$$

необозримо длинна, но мы его определили достаточно просто. Надо подвергнуть сравнительному изучению различные способы задания объектов из $\mathcal{D}$. Достаточно ограничиться способами задания, которые каждому двоично записанному числу p ставят в соответствие некоторый номер

$$n = S(p)$$

Т.о. способ задания объекта из $\mathcal{D}$ становится ни чем иным, как функцией S от натурального аргумента с натуральными значениями. Немного далее мы обратимся к случаю, когда эта функция вычислима. Такие методы задания можно назвать "эффективными". Но пока сохраним полную общность. Для каждого объекта из $\mathcal{D}$ естественно рассмотреть приводящие к нему p наименьшей длины $\ell(p)$. Эта наименьшая длина и будет "сложностью" объекта

-7-

x при "способе задания S":

$$K_S(x) = \min_{S(p)=n(x)} \ell(p).$$

На языке вычислительной математики можно называть p "программой", а S – "методом программирования". Тогда можно будет сказать, что $K_S(x)$ есть минимальная длина программы, по которой можно получить объект x при методе программирования S.

Если имеется несколько различных способов задания элементов из $\mathcal{D}$

$$S_1, S_2, \ldots, S_r,$$

то легко построить новый метод S, который будет доставлять нам любой объект $x \in \mathcal{D}$ со сложностью $K_S(x)$ лишь примерно на $\log r$ превосходящей минимум из сложностей

$$K_{S_1}(x), K_{S_2}(x), \ldots, K_{S_r}(x).$$

Построение такого метода очень просто. Надо резервировать достаточное число начальных знаков последовательности p для фиксации метода S_i, которому надо следовать, пользуясь как программой оставшимися знаками p.

Скажем, что метод S "с точностью до ℓ поглощает метод S'", если всегда

$$K_S(x) \leq K_{S'}(x) + \ell$$

Выше показано, как построить метод S, который с точностью до ℓ сильнее любого из методов $S_1, S_2, \ldots, S_r$, где приблизительно $\ell \sim \log r$.

Randomness for finite objects

- One of these two sequences was obtained by a **physical random process**:
  - 1010011100101011101101110010100001100 01
  - 10101010101010101010101010101010101010 1

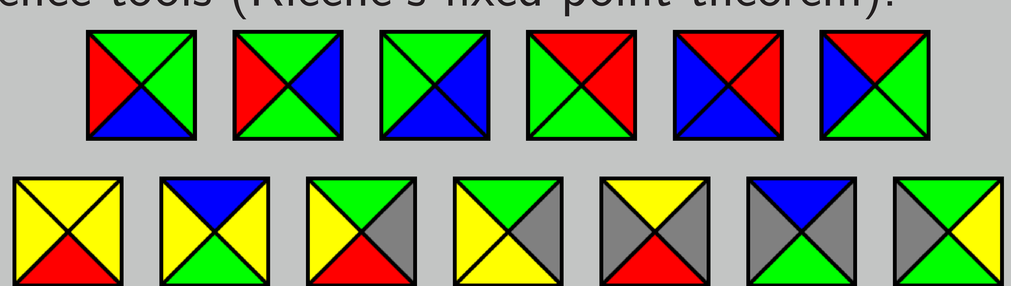
  Which one? Can you justify your answer?
- **Algorithmic information theory**: *Randomness* is high **complexity** (the shortest program that prints a sequence is as long as the sequence itself)
- For finite objects: **randomness = maximal complexity**

Randomness for infinite sequences

- **Martin-Löf**: infinite sequences may be *random* or *non-random*
- **Non-random sequence** = a sequence that violates some effective law of probability theory (statement that is true for all sequences except for some *effectively* null set)
- **Schnorr characterization**: sequence is random *iff* no martingale (gambling system) wins infinitely much against it
- **Levin–Schnorr characterization**: sequence is random *iff* its prefixes are incompressible

Local rules and global complexity

- **Nature**: local interaction in crystals produces nice periodic structures we observe; we believe that local interaction produces nice aperiodic structures in quasicrystals
- **Mathematical model of local interaction**: tiles
- Many tile sets are known that produce **aperiodic** tilings
- A technique for constructing **robust tilings** is developed based on computer science tools (Kleene's fixed-point theorem).



Randomness paradox

A factory produces decks of cards. To make them ready for use, after the printing machine there is a shuffling machine that puts cards in a random order. The management wants to add a quality control unit that checks whether the shuffling machine does its job correctly.

Should the **quality control** reject some decks of cards as '*badly shuffled*'?

- yes, to prevent an angry customer saying that he bought a deck and found the cards in an increasing order;
- no: rejecting some orders destroys the randomness which requires that all ordering appear equally often.

On-line randomness

- Randomness depends on **context**
- *Example*: each match in a football tournament starts with coin tossing (it determines who starts the game)
- bad: if the outcome of today's coin tossing can be computed from the contents of yesterday's newspaper
- normal: if it can be computed from the tomorrow's newspaper
- mathematical **definition** of on-line randomness and on-line complexity; generalization of **Levin–Schnorr characterization** of randomness for the online case; generalization of **martingale characterization** of randomness for the online case; relation to randomness with respect to the class of measures; **non-additivity** for on-line complexity (with possible application to the causality problem).

Online randomness: formal definition

- *Formally*: $\mathbf{x_1, b_1, x_2, b_2, x_3, b_3, \ldots}$; here $\mathbf{x_i}$ are strings representing context information, $\mathbf{b_i}$ are presumable random bits
- *Example*: $\mathbf{b_i = \Phi(x_i)}$ is bad ... and $\mathbf{b_i = \Phi(x_{i+1})}$ is OK
- **On-line measure**: No probability assumption on $\mathbf{x_i}$; assumed conditional probabilities for $\mathbf{b_i}$
- **On-line effectively null sets**: for every $\varepsilon > 0$ one can algorithmically find a covering by interval with *upper* probability less than ε
- Upper probability can be defined as the upper bound of probabilities *wrt* all distributions where conditional probability of each next bit is $1/2$

On-line complexity

- **complexity** of $(\mathbf{x_1 \to b_1; x_2 \to b_2; \ldots})$ = the minimal length of an *on-line* program that gets $\mathbf{x_1}$, prints $\mathbf{b_1}$, then gets $\mathbf{x_2}$, prints $\mathbf{b_2}$, *etc.*
- **Theorem**: in the sequence of bits $\mathbf{x_1, y_1, x_2, y_2, \ldots}$

  $$\mathbf{K(x_1 \to y_1; x_2 \to y_2; \ldots) + K(x_1; y_1 \to x_2; y_2 \to x_3; \ldots)}$$

  cannot be smaller than the complexity of $\mathbf{K(x_1y_1x_2y_2\ldots)}$ but can significantly exceed the latter, and also $\mathbf{K(y_1 \to x_1; y_2 \to x_2, \ldots)}$.
- Philosophers can interpret these results as a mathematical way to reconstruct causality from dependence.

An object that combines structure and randomness