Information Theory and Statistics
Lecture 8: Complexity and entropy

Łukasz Dębowski
ldebowsk@ipipan.waw.pl

Ph. D. Programme 2013/2014

Project co-financed by the European Union
within the framework of the European Social Fund

Kolmogorov complexity and entropy

- Prefix-free complexity is the length of a prefix-free code.
- Hence we may expect that it can be related to entropy.

Project co-financed by the European Union
within the framework of the European Social Fund
KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI
PhD STUDIES
UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Distribution of random variables

- Let $\mathbf{(X_i)_{i\in\mathbb{N}}}$ be a sequence of random variables, $\mathbf{X_i : \Omega \to \Gamma}$.
- We denote its a probability distribution

$$\mathbf{P(x_1^m) = P(X_1^m = x_1^m).}$$

- We will consider Kolmogorov complexity $\mathbf{K(x_1^m|P)}$, which is the prefix-free Kolmogorov complexity of $\mathbf{x_1^m}$ given the definition of distribution of $\mathbf{P}$ on the infinite tape.
- If $\mathbf{P}$ is computable then

$$\mathbf{K(x_1^m|P) \stackrel{+}{<} K(x_1^m) \stackrel{+}{<} K(x_1^m|P) + K(P).}$$

Shannon-Fano coding

Theorem

For any distribution $\mathbf{P}$,

$$\mathbf{K}(\mathbf{x}_1^m|\mathbf{P}) \overset{+}{<} -\log \mathbf{P}(\mathbf{x}_1^m) + 2\log \mathbf{m}.$$

Proof

The inequality follows from the fact that a certain program that computes $\mathbf{x}_1^m$ has form "having the definition of $\mathbf{P}$ and the length of string $\mathbf{x}_1^m$, take the Shannon-Fano code word for $\mathbf{x}_1^m$ with respect to $\mathbf{P}$ and compute $\mathbf{x}_1^m$ from it."

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Source coding inequality

Theorem (source coding inequality)

Let $\mathbf{B} : \Gamma^m \to \Gamma^*$ *be a prefix-free code. For any distribution* $\mathbf{P}$,

$$\sum_{x_1^m} P(x_1^m)\,[|B(x_1^m)| + \log P(x_1^m)] \geq 0.$$

Since prefix-free Kolmogorov complexity is the length of a prefix-free code, we obtain the following result:

Theorem

$$0 \leq \sum_{x_1^m} P(x_1^m)\,[K(x_1^m|P) + \log P(x_1^m)] \overset{+}{<} 2\log m.$$

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Barron theorem

Theorem (Barron theorem)

Let $\mathbf{B} : \Gamma^* \to \Gamma^*$ *be a prefix-free code. For any distribution* $\mathbf{P}$,

$$\lim_{m\to\infty} [|B(X_1^m)| + \log P(X_1^m)] = \infty$$

holds with $\mathbf{P}$*-probability* $\mathbf{1}$.

Since prefix-free Kolmogorov complexity is the length of a prefix-free code, we obtain the following result:

Theorem

$$0 \leq [K(X_1^m|P) + \log P(X_1^m)] \overset{+}{<} 2\log m$$

holds for sufficiently large $\mathbf{m}$ *with* $\mathbf{P}$*-probability* $\mathbf{1}$.

Project co-financed by the European Union
within the framework of the European Social Fund
KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI
PhD STUDIES
UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Markov inequality

Theorem (Markov inequality)

Let $\epsilon > \mathbf{0}$ be a fixed constant and let $\mathbf{Y}$ be a random variable such that $\mathbf{Y} \geq \mathbf{0}$. We have

$$\mathbf{P}(\mathbf{Y} \geq \epsilon) \leq \frac{\mathbf{E}\,\mathbf{Y}}{\epsilon}.$$

Proof

Consider random variable $\mathbf{Z} = \mathbf{Y}/\epsilon$. We have

$$\mathbf{P}(\mathbf{Y} \geq \epsilon) = \int_{\mathbf{Z} \geq 1} d\mathbf{P} \leq \int_{\mathbf{Z} \geq 1} \mathbf{Z} d\mathbf{P} \leq \int \mathbf{Z} d\mathbf{P} = \frac{\mathbf{E}\,\mathbf{Y}}{\epsilon}.$$

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Borel-Cantelli lemma

Denote

$$\limsup_{n\to\infty} A_n = \{\omega : \omega \in A_m \text{ for infinitely many } m\}\,.$$

We have

$$\left(\limsup_{n\to\infty} A_n\right)^c = \{\omega : \omega \notin A_m \text{ for sufficiently large } m\}\,.$$

For proving that some events hold with probability 1, the following proposition is particularly useful.

Theorem (Borel-Cantelli lemma)

If $\sum_{m=1}^{\infty} P(A_m) < \infty$ *for a family of events* $A_1, A_2, A_3, \ldots$ *then*

$$P\left(\limsup_{n\to\infty} A_n\right) = 0.$$

Project co-financed by the European Union
within the framework of the European Social Fund

Proof of the Borel-Cantelli lemma

Notice that $\sum_{m=1}^{\infty} P(A_m) < \infty$ implies

$$\lim_{m\to\infty} \sum_{k=m}^{\infty} P(A_k) = 0.$$

Hence we obtain

$$\begin{aligned}
&P(\{\omega : \omega \in A_m \text{ for infinitely many } m\}) \\
&= P\left(\left\{\omega : \forall_{m\geq 1}\exists_{k\geq m}\, \omega \in A_k\right\}\right) \\
&= P\left(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k\right) \\
&\leq \inf_{m\geq 1} P\left(\bigcup_{k=m}^{\infty} A_k\right) \leq \inf_{m\geq 1} \sum_{k=m}^{\infty} P(A_k) = 0.
\end{aligned}$$

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Proof of Barron theorem

Let us write

$$W(x_1^m) = \frac{2^{-|B(x_1^m)|}}{P(x_1^m)2^{-n}}.$$

By the Markov inequality we obtain

$$\begin{aligned}
&\sum_{m=1}^{\infty} P\left(|B(X_1^m)| + \log P(X_1^m) \le n\right) \\
&= \sum_{m=1}^{\infty} P\left(W(X_1^m) \ge 1\right) \\
&\le \sum_{m=1}^{\infty} \sum_{x_1^m} P(x_1^m) W(x_1^m) = \sum_{m=1}^{\infty} \sum_{x_1^m} 2^{-|B(x_1^m)|+n}
\end{aligned}$$

Project co-financed by the European Union
within the framework of the European Social Fund

Proof (continued)

Continuing, by the Kraft inequality we obtain,

$$\sum_{m=1}^{\infty} P\left(|B(X_1^m)| + \log P(X_1^m) \leq n\right)$$

$$\leq \sum_{m=1}^{\infty} \sum_{x_1^m} 2^{-|B(x_1^m)|+n} \leq 2^n < \infty.$$

Hence from the Borel-Cantelli lemma we obtain that

$$|B(X_1^m)| + \log P(X_1^m) > n \text{ for sufficiently large } m$$

holds with P-probability 1.

The n in this statement is arbitrary so the claim follows.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

An analogue of the chain rule

The parallels between prefix-free complexity and entropy can be drawn further. The following theorem is an analogue of the chain rule $\mathbf{H(X, Y) = H(X) + H(Y|X)}$.

Theorem

$$\mathbf{K(\langle u, w\rangle) \stackrel{+}{=} K(u) + K(w|\,\langle u, K(u)\rangle).}$$

In the proposition above, it is easy to show that the left hand side is smaller than the right hand side. The proof of the converse inequality is harder.

Project co-financed by the European Union
within the framework of the European Social Fund

Partial proof of the chain rule

We will demonstrate that

$$\mathbf{K}(\langle \mathbf{u}, \mathbf{w} \rangle) \stackrel{+}{<} \mathbf{K}(\mathbf{u}) + \mathbf{K}(\mathbf{w} | \langle \mathbf{u}, \mathbf{K}(\mathbf{u}) \rangle).$$

Let $\mathbf{p}$ be the shortest program that satisfies $\mathbf{V}(\mathbf{p}) = \mathbf{u}$ and let $\mathbf{p}'$ be the shortest program that satisfies $\mathbf{V}(\mathbf{p}' | \langle \mathbf{u}, \mathbf{K}(\mathbf{u}) \rangle) = \mathbf{w}$. Then there exists a prefix-free machine $\mathbf{S}$ that satisfies $\mathbf{S}(\mathbf{p}\mathbf{p}') = \langle \mathbf{u}, \mathbf{w} \rangle$. Hence we obtain the claim.

Project co-financed by the European Union
within the framework of the European Social Fund

Incomplete analogy

In the algorithmic chain rule there appears term $\mathbf{K}(\mathbf{w}|\langle\mathbf{u},\mathbf{K}(\mathbf{u})\rangle)$ rather than $\mathbf{K}(\mathbf{w}|\mathbf{u})$. Although $\mathbf{K}(\mathbf{w}|\langle\mathbf{u},\mathbf{K}(\mathbf{u})\rangle)$ differs from $\mathbf{K}(\mathbf{w}|\mathbf{u})$, we can see that $\mathbf{K}(\langle\mathbf{u},\mathbf{K}(\mathbf{u})\rangle)$ and $\mathbf{K}(\mathbf{u})$ are approximately equal.

Theorem

$$\mathbf{K}(\langle\mathbf{w},\mathbf{K}(\mathbf{w})\rangle) \stackrel{+}{=} \mathbf{K}(\mathbf{w}).$$

Proof

From the shortest program that computes $\mathbf{w}$, we may reconstruct both $\mathbf{w}$ and $\mathbf{K}(\mathbf{w})$. Hence $\mathbf{K}(\langle\mathbf{w},\mathbf{K}(\mathbf{w})\rangle) \stackrel{+}{<} \mathbf{K}(\mathbf{w})$. On the hand, we have $\mathbf{K}(\langle\mathbf{w},\mathbf{K}(\mathbf{w})\rangle) \stackrel{+}{>} \mathbf{K}(\mathbf{w})$ from a previous theorem.

Project co-financed by the European Union
within the framework of the European Social Fund

Algorithmic information

Definition

We define *algorithmic information* between strings $\mathbf{u}$ and $\mathbf{w}$ as

$$\mathbf{I}(\mathbf{u};\mathbf{w}) = \mathbf{K}(\mathbf{w}) - \mathbf{K}(\mathbf{w}|\ \langle \mathbf{u}, \mathbf{K}(\mathbf{u})\rangle).$$

Project co-financed by the European Union
within the framework of the European Social Fund

Symmetry of algorithmic information

Theorem

$$\mathbf{I}(\mathbf{u};\mathbf{w}) \stackrel{+}{=} \mathbf{I}(\mathbf{w};\mathbf{u}).$$

Proof

Observe that

$$\begin{aligned} \mathbf{I}(\mathbf{u};\mathbf{w}) &= \mathbf{K}(\mathbf{w}) - \mathbf{K}(\mathbf{w}|\langle \mathbf{u}, \mathbf{K}(\mathbf{u})\rangle) \\ &\stackrel{+}{=} \mathbf{K}(\mathbf{w}) + \mathbf{K}(\mathbf{u}) - \mathbf{K}(\langle \mathbf{u}, \mathbf{w}\rangle) \\ &\stackrel{+}{=} \mathbf{K}(\mathbf{u}) - \mathbf{K}(\mathbf{u}|\langle \mathbf{w}, \mathbf{K}(\mathbf{w})\rangle) = \mathbf{I}(\mathbf{w};\mathbf{u}). \end{aligned}$$

Project co-financed by the European Union
within the framework of the European Social Fund