Hilberg Exponents: New Measures of Long Memory in the Process

Łukasz Dębowski
ldebowsk@ipipan.waw.pl

Institute of Computer Science
Polish Academy of Sciences

12th November 2015, Gifu

1 Introduction

2 Relating $\gamma_{\mathbf{P}}^{\pm}$ and $\delta_{\mathbf{P}}^{\pm}$

3 Relating $\delta_{\mathbf{P}}^{+}$ and $\delta_{\mathbf{R}}^{+}$

4 Evaluating $\gamma_{\mathbf{Q}}^{\pm}$ and $\delta_{\mathbf{Q}}^{\pm}$

5 Conclusion

Motivation: Hilberg's hypothesis

- According to a hypothesis by Hilberg (1990), the mutual information between two adjacent blocks of text in natural language grows like a power of the block length.
- This property differentiates natural language from **k**-parameter sources, for which the mutual information is proportional to the logarithm of the block length.
- In 2011, we constructed processes, called Santa Fe processes, which feature the power-law growth of mutual information.
- In 2011, we also showed that Hilberg's hypothesis implies Herdan's law, some version of Zipf's law.
- In 2014, we showed experimentally that for a PPM-like code the estimates of mutual information grow as a power law for natural language and logarithmically for a **k**-parameter source.

Preliminaries

1. $\mathbb{X}$ — a countable alphabet,
   $(\Omega, \mathcal{J}, Q)$ — probability space with $\Omega = \mathbb{X}^{\mathbb{Z}}$,
   $X_k : \Omega \ni (x_i)_{i\in\mathbb{Z}} \mapsto x_k \in \mathbb{X}$ — random variables,
   Q — a stationary measure (not necessarily ergodic),
   $X_n^m = (X_i)_{n\le i\le m}$ — blocks of symbols,
   $E_Q X$ — expectation,
   $\mathrm{Var}_Q X$ — variance.
2. P — a code (incomplete measure), i.e., it satisfies $P(x_1^n) \ge 0$ and the Kraft inequality $\sum_{x_1^n} P(x_1^n) \le 1$.
   For P (or for Q), we define the pointwise mutual information

   $$I^P(n) = -\log P(X_{-n+1}^0) - \log P(X_1^n) + \log P(X_{-n+1}^n).$$

   In the formula $\log$ stands for the binary logarithm.

Hilberg exponents

Define the positive logarithm

$$\log^+ x = \begin{cases} \log(x+1), & x \geq 0, \\ 0, & x < 0. \end{cases}$$

For a code **P** we introduce

$$\gamma_P^+ = \limsup_{n\to\infty} \frac{\log^+ I^P(n)}{\log n}, \qquad \gamma_P^- = \liminf_{n\to\infty} \frac{\log^+ I^P(n)}{\log n},$$

$$\delta_P^+ = \limsup_{n\to\infty} \frac{\log^+ E_Q I^P(n)}{\log n}, \qquad \delta_P^- = \liminf_{n\to\infty} \frac{\log^+ E_Q I^P(n)}{\log n}.$$

We call these: γ_P^+—the upper random Hilberg exponent, γ_P^-—the lower random Hilberg exponent, δ_P^+—the upper expected Hilberg exponent, and δ_P^-—the lower expected Hilberg exponent.

Basic observations

By definition,

$$\gamma_{\mathbf{P}}^{+} \geq \gamma_{\mathbf{P}}^{-} \geq \mathbf{0},$$
$$\delta_{\mathbf{P}}^{+} \geq \delta_{\mathbf{P}}^{-} \geq \mathbf{0}.$$

- For $\mathbf{P} = \mathbf{Q}$, Hilberg exponents quantify some sort of long-range non-Markovian dependence in the process.
- For an IID process or a hidden Markov process with a finite number of hidden states, $\mathbf{E}_{\mathbf{Q}}\, \mathbf{I}^{\mathbf{Q}}(\mathbf{n}) \leq \mathbf{D}$ so $\delta_{\mathbf{Q}}^{\pm} = \mathbf{0}$.
- For a $\mathbf{k}$-parameter source, $\mathbf{E}_{\mathbf{Q}}\, \mathbf{I}^{\mathbf{Q}}(\mathbf{n}) \propto \mathbf{k} \log \mathbf{n}$, so $\delta_{\mathbf{Q}}^{\pm} = \mathbf{0}$.
- If $\mathbf{E}_{\mathbf{Q}}\, \mathbf{I}^{\mathbf{Q}}(\mathbf{n}) \propto \mathbf{n}^{\beta}$ where $\beta \in [\mathbf{0}, \mathbf{1}]$ then $\delta_{\mathbf{Q}}^{\pm} = \beta$.
- There exist some non-Markovian but mixing sources, being a generalization of Santa Fe processes, for which $\delta_{\mathbf{Q}}^{\pm} \in (\mathbf{0}, \mathbf{1})$.

Further simple observations

Inequalities

$$\gamma_P^- \le \gamma_P^+ \le 1,$$
$$\delta_P^- \le \delta_P^+ \le 1$$

hold in the following cases:

1. For $P = Q$ — by the Shannon-McMillan-Breiman theorem and by stationarity.
2. For P being universal almost surely and in expectation, that is if for $k(n)$ and $l(n)$ being nondecreasing functions of n, where $k(n) + l(n) \to \infty$ we have

$$\lim_{n\to\infty} \frac{1}{k(n)+l(n)+1}\left[-\log P(X_{-k(n)}^{l(n)})\right] = h_Q,$$

where h_Q is the entropy rate of measure Q, and if

$$\lim_{n\to\infty} \frac{1}{k(n)+l(n)+1} E_Q\left[-\log P(X_{-k(n)}^{l(n)})\right] = E_Q\, h_Q.$$

A research program

There are many Hilberg exponents, for different measures and for different codes. Seeking for some order, we may look for results of three kinds:

1. For a fixed code $\mathbf{P}$ and a measure $\mathbf{Q}$, we relate the random exponents $\gamma_{\mathbf{P}}^{\pm}$ and the expected exponents $\delta_{\mathbf{P}}^{\pm}$.
2. For two codes $\mathbf{P}$ and $\mathbf{R}$, we relate the exponents of a fixed kind, say $\delta_{\mathbf{P}}^{+}$ and $\delta_{\mathbf{R}}^{+}$ for some measure $\mathbf{Q}$.
3. For a fixed code $\mathbf{P}$ and a measure $\mathbf{Q}$, we directly evaluate exponents $\gamma_{\mathbf{P}}^{\pm}$ and $\delta_{\mathbf{P}}^{\pm}$.

In the following we will present some results of these three sorts.

1. Introduction
2. Relating $\gamma_{\mathbf{P}}^{\pm}$ and $\delta_{\mathbf{P}}^{\pm}$
3. Relating $\delta_{\mathbf{P}}^{+}$ and $\delta_{\mathbf{R}}^{+}$
4. Evaluating $\gamma_{\mathbf{Q}}^{\pm}$ and $\delta_{\mathbf{Q}}^{\pm}$
5. Conclusion

"Second-order" Shannon-McMillan-Breiman theorem

- The original idea of the SMB theorem was to relate the asymptotic growth of pointwise and expected entropies for an ergodic process $\mathbf{Q}$ with $\mathbf{P} = \mathbf{Q}$.
- In contrast, relating the random Hilberg exponents $\gamma_{\mathbf{Q}}^{\pm}$ and the expected Hilberg exponents $\delta_{\mathbf{Q}}^{\pm}$ means relating the speed of growth of the pointwise and expected mutual informations, which is a subtler effect than the SMB theorem.
- Thus relating $\gamma_{\mathbf{Q}}^{\pm}$ and $\delta_{\mathbf{Q}}^{\pm}$ could be called a "second-order" analogue of the SMB theorem.

Our main result

For a code $\mathbf{P}$ with exponent $\delta_{\mathbf{P}}^- > 0$, let us introduce

$$\epsilon_{\mathbf{P}} = \limsup_{n\to\infty} \frac{\log^+ \left[\operatorname{Var}_Q I^P(n) / \operatorname{E}_Q I^P(n)\right]}{\log n}.$$

Theorem

For an ergodic measure $\mathbf{Q}$ over a finite alphabet, random Hilberg exponents $\gamma_{\mathbf{Q}}^{\pm}$ are almost surely constant. Moreover, we have $\mathbf{Q}$-almost surely

$$\begin{aligned} \delta_{\mathbf{Q}}^+ &\geq \gamma_{\mathbf{Q}}^+ \geq \delta_{\mathbf{Q}}^+ - \epsilon_{\mathbf{Q}}, \\ \delta_{\mathbf{Q}}^- &\geq \gamma_{\mathbf{Q}}^- \geq \delta_{\mathbf{Q}}^- - \epsilon_{\mathbf{Q}}, \end{aligned}$$

where the left inequalities hold without restrictions, whereas the right inequalities hold for $\delta_{\mathbf{Q}}^- > 0$.

The fundamental idea of the proof

Our theorem can be demonstrated without invoking the ergodic theorem. Instead, we use an auxiliary "Kolmogorov code"

$$\mathbf{S}(\mathbf{x}_1^n) = 2^{-\mathbf{K}(\mathbf{x}_1^n|\mathbf{F})},$$

where $\mathbf{K}(\mathbf{x}_1^n|\mathbf{F})$ is the prefix-free Kolmogorov complexity of a string $\mathbf{x}_1^n$ given an object $\mathbf{F}$ on an additional infinite tape. The object $\mathbf{F}$ can be another string or, here, a definition of measure $\mathbf{Q}$.

The first auxiliary result

Theorem

Consider Kolmogorov code $\mathbf{S}$ *and an ergodic* $\mathbf{Q}$ *over a finite alphabet* $\mathbb{X}$*. Exponents* $\gamma_{\mathbf{S}}^{-}$ *and* $\gamma_{\mathbf{S}}^{+}$ *are* $\mathbf{Q}$*-almost surely constant.*

The idea of the proof:

1. $\left|\mathbf{K}(\mathbf{x}_1^n|\mathbf{F}) - \mathbf{K}(\mathbf{x}_{t+1}^{t+n}|\mathbf{F})\right| \le \mathbf{C}t.$
2. Hence $\gamma_{\mathbf{S}}^{\pm}$ are shift invariant.
3. Hence $\gamma_{\mathbf{S}}^{\pm}$ are constant on ergodic sources.

Second auxiliary result (via Borel-Cantelli lemma)

Consider Kolmogorov code $\mathbf{S}$. For $\mathbf{I}^{\mathbf{S}}(\mathbf{n}) + \mathbf{B} \geq \mathbf{1}$, define

$$\zeta_{\mathbf{S}}^{+} = \limsup_{\mathbf{n}\to\infty} \frac{\log^{+}\left[\mathbf{E}_{\mathbf{Q}}(\mathbf{I}^{\mathbf{S}}(\mathbf{n}) + \mathbf{B})^{-1}\right]^{-1}}{\log \mathbf{n}},$$

$$\zeta_{\mathbf{S}}^{-} = \liminf_{\mathbf{n}\to\infty} \frac{\log^{+}\left[\mathbf{E}_{\mathbf{Q}}(\mathbf{I}^{\mathbf{S}}(\mathbf{n}) + \mathbf{B})^{-1}\right]^{-1}}{\log \mathbf{n}}.$$

These will be called inverse expected Hilberg exponents.

Theorem

Consider Kolmogorov code $\mathbf{S}$ *and a stationary measure* $\mathbf{Q}$*. Then:*

1. $\delta_{\mathbf{S}}^{+} \geq \gamma_{\mathbf{S}}^{+}$ $\mathbf{Q}$*-almost surely and* $\operatorname{ess\,sup}_{\mathbf{Q}} \gamma_{\mathbf{S}}^{+} \geq \zeta_{\mathbf{S}}^{+}$.
2. $\delta_{\mathbf{S}}^{-} \geq \operatorname{ess\,inf}_{\mathbf{Q}} \gamma_{\mathbf{S}}^{-}$ *and* $\gamma_{\mathbf{S}}^{-} \geq \zeta_{\mathbf{S}}^{-}$ $\mathbf{Q}$*-almost surely.*

A corollary

Corollary

For an ergodic measure $\mathbf{Q}$ *over a finite alphabet, equalities* $\gamma_{\mathbf{S}}^{+} = \operatorname{\mathbf{ess\,sup}}_{\mathbf{Q}} \gamma_{\mathbf{S}}^{+}$ *and* $\gamma_{\mathbf{S}}^{-} = \operatorname{\mathbf{ess\,inf}}_{\mathbf{Q}} \gamma_{\mathbf{S}}^{-}$ *hold* $\mathbf{Q}$*-almost surely. Hence,* $\mathbf{Q}$*-almost surely we have*

$$\delta_{\mathbf{S}}^{+} \geq \gamma_{\mathbf{S}}^{+} \geq \zeta_{\mathbf{S}}^{+},$$
$$\delta_{\mathbf{S}}^{-} \geq \gamma_{\mathbf{S}}^{-} \geq \zeta_{\mathbf{S}}^{-}.$$

Third auxiliary result

Theorem

Consider Kolmogorov code $\mathbf{S}$ *with* $\mathbf{F} = \mathbf{Q}$, *where* $\mathbf{Q}$ *is a stationary measure. Then:*

1. $\delta_{\mathbf{S}}^{-} = \delta_{\mathbf{Q}}^{-}$ *and* $\delta_{\mathbf{S}}^{+} = \delta_{\mathbf{Q}}^{+}$.
2. $\gamma_{\mathbf{S}}^{-} = \gamma_{\mathbf{Q}}^{-}$ *and* $\gamma_{\mathbf{S}}^{+} = \gamma_{\mathbf{Q}}^{+}$ $\mathbf{Q}$*-almost surely.*

(By Shannon-Fano coding and Barron's inequality.)

Fourth, the last auxiliary result

For a code $\mathbf{P}$ with exponent $\delta_{\mathbf{P}}^{-} > \mathbf{0}$, let us introduce

$$\epsilon_{\mathbf{P}} = \limsup_{\mathbf{n}\to\infty} \frac{\log^{+}\left[\mathrm{Var}_{\mathbf{Q}}\, \mathbf{I}^{\mathbf{P}}(\mathbf{n}) / \mathbf{E}_{\mathbf{Q}}\, \mathbf{I}^{\mathbf{P}}(\mathbf{n})\right]}{\log \mathbf{n}}.$$

Theorem

Consider Kolmogorov code $\mathbf{S}$ and a stationary measure $\mathbf{Q}$. If $\delta_{\mathbf{S}}^{-} > \mathbf{0}$ then $\zeta_{\mathbf{S}}^{+} \geq \delta_{\mathbf{S}}^{+} - \epsilon_{\mathbf{S}}$ and $\zeta_{\mathbf{S}}^{-} \geq \delta_{\mathbf{S}}^{-} - \epsilon_{\mathbf{S}}$.

(By Markov inequality.)

Theorem

Consider Kolmogorov code $\mathbf{S}$ with $\mathbf{F} = \mathbf{Q}$, where $\mathbf{Q}$ is a stationary measure. If $\delta_{\mathbf{Q}}^{-} > \mathbf{0}$ then $\epsilon_{\mathbf{S}} = \epsilon_{\mathbf{Q}}$.

(By Shannon-Fano coding and Barron's inequality.)

Resuming, our main result

Theorem

For an ergodic measure $\mathbf{Q}$ *over a finite alphabet, random Hilberg exponents* $\gamma_{\mathbf{Q}}^{\pm}$ *are almost surely constant. Moreover, we have* $\mathbf{Q}$*-almost surely*

$$\delta_{\mathbf{Q}}^{+} \ge \gamma_{\mathbf{Q}}^{+} \ge \delta_{\mathbf{Q}}^{+} - \epsilon_{\mathbf{Q}},$$
$$\delta_{\mathbf{Q}}^{-} \ge \gamma_{\mathbf{Q}}^{-} \ge \delta_{\mathbf{Q}}^{-} - \epsilon_{\mathbf{Q}},$$

where the left inequalities hold without restrictions, whereas the right inequalities hold for $\delta_{\mathbf{Q}}^{-} > 0$.

1. Introduction
2. Relating $\gamma_{\mathbf{P}}^{\pm}$ and $\delta_{\mathbf{P}}^{\pm}$
3. Relating $\delta_{\mathbf{P}}^{+}$ and $\delta_{\mathbf{R}}^{+}$
4. Evaluating $\gamma_{\mathbf{Q}}^{\pm}$ and $\delta_{\mathbf{Q}}^{\pm}$
5. Conclusion

An alternative expression for δ_P^+

Denote $H^P(n) := -\log P(X_1^n)$. Suppose that

$$\lim_{n\to\infty} \frac{1}{n} E_Q H^P(n) = E_Q h_Q$$

and suppose that $E_Q I^P(n) \ge -D$ for a certain $D > 0$. Then

$$\delta_P^+ = \limsup_{n\to\infty} \frac{\log^+ E_Q I^P(n)}{\log n} = \limsup_{n\to\infty} \frac{\log^+ E_Q \left[H^P(n) - h_Q n\right]}{\log n}$$

follows from the telescope sum

$$E_Q H^P(n) - hn = \sum_{k=0}^{\infty} \frac{E_Q I^P(2^k n)}{2^{k+1}}.$$

A hierarchy of approximations of block entropy

1. $R(X_1^n) = 2^{-|C(X_1^n)|}$ — a computable universal code.
2. $P(X_1^n) = 2^{-K(X_1^n)}$ — the unconditional Kolmogorov code.
3. $Q(X_1^n)$ — the underlying measure.
4. $E(X_1^n) = Q(X_1^n|\mathcal{I})$ — the random ergodic measure.
5. $H^T(n) = H^T(n; X_1^{n(|\mathbb{X}|+\epsilon)^n})$ — the plugin entropy estimator.

We have

$$\mathbf{E}_Q\, H^R(n) \geq \mathbf{E}_Q\, H^P(n) \geq \mathbf{E}_Q\, H^Q(n) \geq \mathbf{E}_Q\, H^E(n) \geq \mathbf{E}_Q\, H^T(n),$$

whereas the common rate of these is $\mathbf{E}_Q\, h_Q$. Hence

$$\delta_R^+ \geq \delta_P^+ \geq \delta_Q^+ \geq \delta_E^+ \geq \delta_T^+.$$

The difference $\delta_P^+ - \delta_E^+$ can be arbitrarily close to 1.

1. Introduction
2. Relating $\gamma_{\mathbf{P}}^{\pm}$ and $\delta_{\mathbf{P}}^{\pm}$
3. Relating $\delta_{\mathbf{P}}^{+}$ and $\delta_{\mathbf{R}}^{+}$
4. Evaluating $\gamma_{\mathbf{Q}}^{\pm}$ and $\delta_{\mathbf{Q}}^{\pm}$
5. Conclusion

Memoryless sources and hidden Markov processes

- For IID processes, $\delta_{\mathbf{Q}}^{\pm} = \mathbf{0}$ and hence $\gamma_{\mathbf{Q}}^{\pm} = \mathbf{0}$ since there is no dependence in the process.
- For Markov processes over a finite alphabet and hidden Markov processes with a finite number of hidden states, we also have $\delta_{\mathbf{Q}}^{\pm} = \mathbf{0}$ and hence $\gamma_{\mathbf{Q}}^{\pm} = \mathbf{0}$, since the expected mutual information is bounded for measures of those processes by the data-processing inequality.

Mixture Bernoulli process

Some simple example of a process with unbounded mutual information is the mixture of Bernoulli processes over the alphabet $\mathbb{X} = \{0, 1\}$, which we will call the mixture Bernoulli process:

$$Q(x_1^n) = \int_0^1 \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} d\theta = \frac{1}{n+1}\binom{n}{\sum_{i=1}^n x_i}^{-1}.$$

Mixture Bernoulli process (continued)

$$T_n = \sum_{i=-n+1}^{0} X_i, \qquad S_n = \sum_{i=1}^{n} X_i.$$

Theorem

For the mixture Bernoulli process, $\delta_Q^{\pm} = \gamma_Q^{\pm} = 0$.

Proof: X_{-n+1}^{0} and X_1^n are independent given T_n and S_n. Hence

$$I^Q(n) = -\log \frac{Q(T_n)Q(S_n)}{Q(T_n, S_n)}$$

so $E_Q\, I^Q(n) = I_Q(T_n; S_n)$. Variable S_n assumes under Q each value in $\{0, 1, ..., n\}$ with equal probability $(n+1)^{-1}$. Hence $0 \le I_Q(T_n; S_n) \le H_Q(S_n) = \log(n+1)$, which implies the claim.

Santa Fe process

The Santa Fe process $(X_i)_{i\in\mathbb{Z}}$ is a sequence of variables

$$X_i = (K_i, Z_{K_i}),$$

where processes $(K_i)_{i\in\mathbb{Z}}$ and $(Z_k)_{k\in\mathbb{N}}$ with

$$\begin{aligned} Q(Z_k = 0) = Q(Z_k = 1) = 1/2, &\qquad (Z_k)_{k\in\mathbb{N}} \sim \text{IID}, \\ Q(K_i = k) = k^{-1/\beta}/\zeta(\beta^{-1}), &\qquad (K_i)_{i\in\mathbb{Z}} \sim \text{IID}, \end{aligned}$$

where $\beta \in (0,1)$ is a parameter and $\zeta(x) = \sum_{k=1}^{\infty} k^{-x}$.

Variable $Y = \sum_{k=1}^{\infty} 2^{-k} Z_k$ could be considered a random real parameter of the process but the distribution of the process $(X_i)_{i\in\mathbb{Z}}$ is not a differentiable function of this parameter.

So, the Santa Fe process is not a 1-parameter source.

Santa Fe process (continued)

Theorem

For the Santa Fe process, $\delta_{\mathbf{Q}}^{\pm} = \gamma_{\mathbf{Q}}^{\pm} = \beta$.

A process with $\delta_Q^+ \neq \delta_Q^-$

Consider a sequence of numbers $(a_k)_{k\in\mathbb{N}}$ where $a_k \in \{0, 1\}$. Let

$$X_i = (K_i, Y_{K_i}),$$

where $Y_k = a_k Z_k$, whereas processes $(K_i)_{i\in\mathbb{Z}}$ and $(Z_k)_{k\in\mathbb{N}}$ are independent and distributed as for the original Santa Fe process.

Theorem

There exists such a sequence $(a_k)_{k\in\mathbb{N}}$ that for the modified Santa Fe process, we have $\delta_Q^+ = \beta$ and $\delta_Q^- = 0$.

1. Introduction
2. Relating $\gamma_{\mathbf{P}}^{\pm}$ and $\delta_{\mathbf{P}}^{\pm}$
3. Relating $\delta_{\mathbf{P}}^{+}$ and $\delta_{\mathbf{R}}^{+}$
4. Evaluating $\gamma_{\mathbf{Q}}^{\pm}$ and $\delta_{\mathbf{Q}}^{\pm}$
5. Conclusion

Conclusion

- We have defined Hilberg exponents — the bounding rates of the power-law growth of mutual information in a process.
- There are surprisingly many meaningful Hilberg exponents, for different measures and different codes.
- We have begun sorting out order in this menagerie but surely there are some interesting hard open problems.

www.ipipan.waw.pl/~ldebowsk