Asymmetry of the Kolmogorov complexity of online predicting odd and even bits

Bruno Bauwens

Université de Lorraine, LORIA, Vandœuvre-lès-Nancy, France
Brbauwens@gmail.com

Abstract

Symmetry of information states that $C(x) + C(y|x) = C(x, y) + O(\log C(x))$. In [3] an online variant of Kolmogorov complexity is introduced and we show that a similar relation does not hold. Let the even (online Kolmogorov) complexity of an n-bitstring $x_1x_2 \dots x_n$ be the length of a shortest program that computes x_2 on input x_1, computes x_4 on input $x_1x_2x_3$, etc; and similar for odd complexity. We show that for all n there exists an n-bit x such that both odd and even complexity are almost as large as the Kolmogorov complexity of the whole string. Moreover, flipping odd and even bits to obtain a sequence $x_2x_1x_4x_3 \dots$, decreases the sum of odd and even complexity to $C(x)$. Our result is related to the problem of inferrence of causality in timeseries.

1998 ACM Subject Classification E.4 Coding and Information Theory

Keywords and phrases (On-line) Kolmogorov complexity, (On-line) Algorithmic Probability, Philosophy of Causality, Information Transfer

Digital Object Identifier 10.4230/LIPIcs.STACS.2014.125

1 Introduction

Imagine two people want to perform a two-person theater play. First suppose that the play consists of only two independent monologues each one performed by one player. Before performing, the players must memorize their part of the play, and the total studying effort for the two players together can be assumed to be equal to the effort for one person to study the whole script.

Now imagine a play consisting of a large dialogue where both players alternate lines. Each player only needs to study their half of the lines, and it is sufficient to remember each line only after hearing the last lines of the other player. Thus each player needs only to remember their incremental amount of information in his lines, and this suggests the total studying effort might be close to the effort for one person to study the whole script.

However, it often happens that after studying only his own lines, an actor can reproduce the whole piece. Sometimes actors just study the whole piece. This suggests that studying each half of the lines can be as hard as studying everything. In other words, the total effort of both players together might be close to twice the effort of studying the full manuscript.

Can we interpret this example in terms of Shannon information theory? In the first case, let a theater play be modeled by a probability density function $P(X, Y)$ where X and Y represent the two monologues. Symmetry of information states that $H(X) + H(Y|X) = H(X, Y)$, i.e. the information in the first part plus the new information in the second part equals the total information. This equality is exact and can be extended to the interactive case where a similar additivity property remains valid, and this contrasts to the story above.

An absolute measure of information in a string is given by its Kolmogorov complexity, which is the minimal length of a program on a universal Turing machine that prints the string.

31st Symposium on Theoretical Aspects of Computer Science (STACS'14).
Editors: Ernst W. Mayr and Natacha Portier; pp. 125–136
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

See section 2 for formal definitions. Symmetry of information for Kolmogorov complexity holds within logarithmic terms [19, 1]: $C(x) + C(y|x) = C(x, y) + O(\log C(x, y))$.

For the interactive case, we need the online variant of Kolmogorov complexity introduced in [3]. Let $C_{\text{ev}}(x)$ denote the length of a shortest program that computes x_2 on input x_1, computes x_4 on input $x_1x_2x_3$, etc.; and similar for $C_{\text{odd}}(x)$. In the above example all x_i with odd i correspond to lines for the first player and the others to the second.

In Theorem 1, we show that there exist infinitely many bitstrings x, such that both $C_{\text{ev}}(x)$ and $C_{\text{odd}}(x)$ are almost as big as $C(x)$, in agreement with our example. In Theorem 2, we show that there exists $c > 0$ such that $(C_{\text{ev}} + C_{\text{odd}} - C)(x) \geq c|x|$, i.e. the online asymmetry of information can be large compared to the length of x. Finally, we raise the question how large $(C_{\text{ev}} + C_{\text{odd}} - C)(x)$ can be in terms of $|x|$. A more direct upper bound is $|x|/2 + O(1)$, and one can raise the question whether this is tight. We show there exists a smaller one: there exists $c > 0$ such that $(C_{\text{ev}} + C_{\text{odd}} - C)(x) \leq (1/2 - c)|x|$ for all large x.

Our main result is stronger and is related to the problem of defining causality in time series. Imagine there exists a complex system (e.g. a brain) and we make some measurements in two parts of it. The measurements are represented by bitstrings x (from some part X of the brain) and y (from some part Y). We perform these measurements regularly and get a sequence of pairs

$$(x_1, y_1), (x_2, y_2), \ldots$$

We assume that both parts are communicating with each other, however, the time resolution is not enough to decide whether y_i is a reply to x_i or vice versa. However, we might compare the *dialogue complexity* $C_{\text{odd}} + C_{\text{ev}}$ of

$$x_1, y_1, x_2, y_2, \ldots$$

and

$$y_1, x_1, y_2, x_2, \ldots$$

and (following Occam's Razor principle) choose an ordering that makes the dialogue complexity minimal. We show that these complexities can differ substantially.

Questions of causality are often raised in neurology and economics. The notions of Granger causality and information transfer reflect the idea of "influence" and our result implies a theoretical notion of asymmetry of influence that does not need to assume a time delay to "transport" information between X and Y in contrast to existing definitions [6, 7, 15, 11].[1]

To understand why (current) practical algorithms need a time delay to make inferences about the direction of influence, consider two variables X, Y with a joint probability density function $P(X, Y)$. Using Shannon entropy, we can quantify the influence of X upon Y as $I(Y; X) = H(Y) - H(Y|X)$. Symmetry of information directly implies that this equals the influence of Y upon X: $H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y)$. In the online setting, mutual information is replaced by information transfer, which is well studied in the engineering literature [4, 15, 10, 14, 18, 11, 13]. For time delays k and $l > k$ the information transfer from $\mathcal{X}$ to $\mathcal{Y}$ is given by

$$H(Y_n|Y_{n-l}, \ldots, Y_{n-1}) - H(Y_n|Y_{n-l}, \ldots, Y_{n-1}, X_{n-l}, \ldots X_{n-k}),$$

[1] In the case of three or more timeseries there exist algorithms that infer directed information flows between some variables in some special cases where enough conditional independence exist among the variables, see [12, p. 19–20, 50]. In our example no independence is assumed.

(if this term is dependent on n, the sum is taken). This quantification of causality coincides with Granger causality [6, 7] if all involved conditional distributions are Gaussian.

If we incorporate a time delay $k \geq 1$, the information transfers from $\mathcal{X}$ to $\mathcal{Y}$ and $\mathcal{Y}$ to $\mathcal{X}$ can be different. On the other hand, for $k = 0$ they are always equal, and this is a corollary of (the conditional version of) symmetry of information. In the offline case, a similar observation holds for algorithmic mutual information: $C(x) - C(x|y) = C(y) - C(y|x) + O(\log C(x, y))$.[2] In the online setting, algorithmic mutual information can be generalized to algorithmic information transfer. For an n-bit x and y the version without time delay is given by

$$IT(x \to y) = C(y) - C_{\text{ev}}(x_1 y_1 \dots x_n y_n)\,.$$

We show that for all $\epsilon > 0$ there are infinitely many pairs (x, y) with $|x| = |y|$ and $C(x, y) \geq \Omega(|x|)$ such that $IT(x \to y) \leq \epsilon C(x, y)$ while $IT(y \to x)$ exceeds $C(x, y) + O(1)$. Hence, in contrast to Shannon information theory, significant online dependence of x_i on y_i might not imply significant online dependence of y_i on x_i.

Warning: The example where influence (and causality) is asymmetric heavily uses that shortest models are not computable. Decompression algorithms used in practice are always total (or can be extended to total ones). On the other hand, if one wants to be practical, it is natural to not only consider total algorithms but algorithms that terminate within some reasonable time bound (say polynomial). On that level non-symmetry may reappear, even for one pair of messages, which was not possible in our setting. For example suppose x_1 represents a pair of large primes and y_1 represents their product, then it is much easier to produce first x_1 and then y_1 then vice versa.

Muchnik paradox is a result about online randomness [9] that is related to our observations. Consider the example from [3]: in a tournament (say chess), a coin toss decides which player starts the next game. Consider the sequence $b_1, w_1, b_2, w_2, \dots$ of coin tosses and winners of subsequent games. This sequence might not be random (the winner might depend on who starts), but we would be surprised if the coin tossing depends on previous winners.

More precisely, a sequence is Martin-Löf random if no lower semicomputable martingale succeeds on it. To define randomness for even bits, we consider martingales that only bet on even bits, i.e. a martingale F satisfies $F(x0) = F(x1)$ if $|x0|$ is odd. The even bits of ω are *online random* if no lower semicomputable martingale succeeds that only bets on even bits. (In our example, coin tosses b_i are unfair if a betting scheme makes us win on $b_1 w_1 b_2 w_2 \dots$ while keeping the capital constant for "bets" on w_i.) In a similar way randomness for odd bits is defined. Muchnik showed that there exists a non-random sequence for which both odd and even bits are online random. Hence, contributed information by the odd and even bits does not "add up". Muchnik's paradox does not hold for the online version of computable randomness (where martingales are restricted to computable ones), and is an artefact of the non-computability of the considered martingales.

The article is organised as follows: the next section presents definitions and results. The subsequent three sections are devoted to the proofs: first theorems are reformulated using online semimeasures, and then lower bounds are proven. In the full version of the paper, which is available on ArXiv, there are four appendices containing: a proof of the chain rule

[2] However, logarithmic deviations can appear, if one considers prefix complexity, for example if y is chosen to be a string consisting of $K(x)$ zeros. In this case, it is known that for each n there exist n-bit x such that $K(K(x)) - K(K(x)|x) \leq O(1)$ while $K(x) - K(x|K(x)) \geq \log n - O(\log \log n)$. Moreover, this small error was exploited in an earlier and more involved proof of Theorem 2 [2].

for online complexity, the generalization of Theorem 1 for online computation with more machines, a version of Theorem 2 with a larger linear constant, and a full proof of the upper bound (Theorem 3).

2 Definitions and results

Kolmogorov complexity of a string x on an optimal machine U is the minimal length of a program that computes x and halts. More precisely, associate with a Turing machine a function U that maps pairs of strings to strings. The conditional Kolmogorov complexity is given by

$$C_U(x|y) = \min\{|p| : U(p,y) = x\} \ .$$

This definition depends on U, but there exist a class of machines for which $C_U(x|y)$ is minimal within an additive constant for all x and y. We fix such an optimal U, and drop this index, see [8, 5] for details. If y is the empty string, we write $C(x)$ in stead of $C(x|y)$, and the complexity of a pair $C(x,y|z)$ is given by applying an injective computable pairing function to x and y.

The *even (online Kolmogorov) complexity* [3] of a string z is

$$C_{\mathrm{ev}}(z) = \min\{|p| : U(p, z_1 \ldots z_{i-1}) = z_i \text{ for all } i = 2, 4, \ldots, \leq |z|\}\,.$$

Again, there exists a class of optimal machines U for which C_{ev} is minimal within an additive constant and we assume that U is such a machine. Note that $C(x|y) - O(1) \leq C_{\mathrm{ev}}(y_1x_1 \ldots y_nx_n) \leq C(x) + O(1)$ for n-bit x and y. Let $C_{\mathrm{ev}}(w|v)$ be the conditional variant. The chain rule for the concatenation vw of strings v and w holds: $C_{\mathrm{ev}}(vw) = C_{\mathrm{ev}}(v) + C_{\mathrm{ev}}(w|v) + O(\log(|v|))$, see the full version of the paper. In a similar way $C_{\mathrm{odd}}(x)$ is defined. A direct lower and upper bound for $C_{\mathrm{odd}} + C_{\mathrm{ev}}$ are[3]

$$C(z) - O(\log|z|) \leq (C_{\mathrm{odd}} + C_{\mathrm{ev}})(z) \leq 2C(z) + O(1)\,.$$

The lower bound is almost tight, for example if all even bits of z are zero. Surprisingly, the upper bound can also be almost tight and $C_{\mathrm{odd}} + C_{\mathrm{ev}}$ can change significantly after a simple permutation of the bits.

▶ **Theorem 1.** *For every $\varepsilon > 0$ there exist $\delta > 0$ and a sequence ω such that for large n*

$$\begin{matrix} C_{\mathrm{odd}}(\omega_1 \ldots \omega_n) \\ C_{\mathrm{ev}}(\omega_1 \ldots \omega_n) \end{matrix} \geq (1-\varepsilon)C(\omega_1 \ldots \omega_n) + \delta n\,.$$

Moreover, for all even n

$$C_{\mathrm{odd}}(\omega_2\omega_1 \ldots \omega_n\omega_{n-1}) = C(\omega_1 \ldots \omega_n) + O(\log n) \qquad (1)$$

$$C_{\mathrm{ev}}(\omega_2\omega_1 \ldots \omega_n\omega_{n-1}) \leq O(1)\,. \qquad (2)$$

The first part implies

$$\limsup_{|x|\to\infty} \frac{C_{\mathrm{odd}}(x) + C_{\mathrm{ev}}(x)}{C(x)} \geq 2\,,$$

[3] The $O(\log|x|)$ term could be decreased to $O(1)$ if we compared online complexity with decision complexity [17] as in [3]. However, plain and decision complexity differ by at most $O(\log|x|)$, and because we focus on linear bounds, we do not use this rare variant of complexity.

and by the upper bound $C_{\text{odd}}, C_{\text{ev}} \leq C + O(1)$, this supremum equals 2. Recall the definition $IT(x \to y) = C(y) - C_{\text{ev}}(x_1 y_1 \dots x_n y_n)$ for x, y, n such that $n = |x| = |y|$. Let $x = \omega_1\omega_3 \dots \omega_{2n-1}$ and $y = \omega_2\omega_4 \dots \omega_{2n}$, Theorem 1 implies

$$\begin{aligned} IT(x \to y) &\leq \varepsilon C(x,y) + O(1) \\ IT(y \to x) &= C(x,y) + O(1)\,, \end{aligned}$$

(where $C(x,y) \geq \delta n - O(1)$).[4]

Theorem 1 can be generalized to dialogues between $k \geq 2$ machines, i.e. if k sources need to perform a dialogue, it can happen that each source must contain almost full information about the dialogue. Moreover, if the order is changed, the "contribution" of all except one source becomes computable. Let the complexity of bits $i \bmod k$ be given by

$$C_{i \bmod k}(x) = \min\left\{|p| : U(p, x_1 \dots x_{j-1}) = x_j \text{ for all } j = i, i+k, \dots, \leq |x|\right\}.$$

For every k and $\varepsilon > 0$ there exist a $\delta > 0$ and a sequence ω such that for all $i \leq k$ and large n

$$C_{i \bmod k}(\omega_1 \dots \omega_n) \geq (1-\varepsilon) C(\omega_1 \dots \omega_n) + \delta n$$

Moreover, for $\tilde{\omega} = \omega_k\omega_1 \dots \omega_{k-1}\,\omega_{2k}\omega_{k+1} \dots \omega_{2k-1}\ \dots$ for all n, and $i = 2 \dots k$:

$$\begin{aligned} C_{1 \bmod k}(\tilde{\omega}_1 \dots \tilde{\omega}_n) &= C(\omega_1 \dots \omega_n) + O(\log n) \\ C_{i \bmod k}(\tilde{\omega}_1 \dots \tilde{\omega}_n) &\leq O(1)\,. \end{aligned}$$

In Theorem 1 the difference between C and $C_{\text{odd}} + C_{\text{ev}}$ is linear in the length of the prefix of ω. One might wonder how big this difference can be. A direct bound is $|x|/2 + O(1)$. Indeed, the odd complexity of x is at most $C(x)$ hence

$$\left(C_{\text{odd}} + C_{\text{ev}}\right)(x) - C(x) = \left(C_{\text{odd}}(x) - C(x)\right) + C_{\text{ev}}(x) \leq O(1) + |x|/2 + O(1)\,.$$

The next theorem shows that the difference can indeed be $c|x|$ for a significant c.

▶ **Theorem 2.** *There exist a sequence ω such that for all n*

$$(C_{\text{odd}} + C_{\text{ev}})(\omega_1 \dots \omega_n) \geq n(\log \tfrac{4}{3})/2 + C(\omega_1 \dots \omega_n) - O(\log n)\,.$$

Moreover, Equations (1) *and* (2) *are satisfied.*

The factor $(\log \frac{4}{3})/2$ can be further improved to $(\log \frac{3}{2})/2 \approx 0.292$ at the cost of weakening (1) and (2) (see full version of this paper). On the other hand, the upper bound $1/2$ can not be reached:

▶ **Theorem 3.** *There exist $\beta < \frac{1}{2}$ such that for large x*

$$\left(C_{\text{ev}} + C_{\text{odd}} - C\right)(x) \leq \beta |x|\,.$$

In summary, $\frac{1}{2}\log\frac{3}{2} \leq \limsup \frac{(C_{\text{ev}} + C_{\text{odd}} - C)(x)}{|x|} < \frac{1}{2}$, but the precise value of the lim sup is unknown.

[4] For the first we use $C(y) \leq C(\omega_{1\dots 2n}) = C(x,y)$ up to $O(1)$ terms. For the second $C(x,y) \geq C(x) \geq C_{\text{ev}}(y_1x_1 \dots y_nx_n) = C(x,y)$, thus $C(x) = C(x,y)$, while $C_{\text{ev}}(y_1x_1 \dots y_nx_n) \leq O(1)$. Also, note that $C(\omega_{1\dots 2n})$ must exceed δn because it exceeds $C_{\text{odd}}(\omega_{1\dots 2n}) \geq \delta n$, all up to $O(1)$ terms.

3 Online semimeasures

We show that the problem of constructing strings where additivity of online complexity is violated is equivalent to constructing lower semicomputable semimeasures that can not be factorized into "odd" and "even" online lower semicomputable semimeasures. Before defining such semimeasures and reformulating Theorems 1–3, we recall the algorithmic coding theorem.

A (continuous) semimeasure P is a function from strings to $[0,1]$ such that $P(x0)+P(x1) \leq P(x)$ for all x. A real function f on strings is lower semicomputable if the set of all pairs (x,r) of strings and rational numbers such that $f(x) \leq r$ is enumerable. There exist a maximal lower semicomputable semimeasure $M(x)$, i.e. a lower semicomputable that exceeds any other such semimeasures within a constant factor: $M(x) = \sum_i 2^{-i} P_i(x)$ for an enumeration P_1, P_2, ... of all such semimeasures (see [5, 8, 16] for details). The coding theorem [8, Theorem 4.3.4] implies

$$\log 1/M(x) = C(x) + O(\log C(x)) \,.$$

An *even (online) semimeasure* [3] is a function from strings to $[0,1]$ such that for all x

i. $P(x0) + P(x1) \leq P(x)$ if $|x0|$ is even,
ii. $P(x0) = P(x1) = P(x)$ otherwise.

The coding theorem generalizes to the online setting.

▶ **Theorem 4** ([3]). *There exist maximal even (respectively odd) semimeasures. All such semimeasures M_{ev} (resp. M_{odd}) satisfy*

$$\log 1/M_{\text{ev}}(x) = C_{\text{ev}}(x) + O\left(\log C_{\text{ev}}(x)\right).$$

Let $\omega_{k\ldots l} = \omega_k \ldots \omega_l$. Theorems 1, 2 and 3 follow from

▶ Proposition 5. For all $\varepsilon > 0$ and lower semicomputable odd and even online semimeasures Q_{odd} and Q_{ev}, there exist δ, a sequence ω, a lower semicomputable semimeasure P, and a partial computable F such that for all n

$$(Q_{\text{odd}}Q_{\text{ev}})(\omega_{1\ldots n}) \leq (1-\delta)^n P(\omega_{1\ldots n})^{2-2\varepsilon}$$

and $F(\omega_{1\ldots 2n}, \omega_{2n+2}) = \omega_{2n+1}$.

▶ Proposition 6. For all lower semicomputable odd and even online semimeasures Q_{odd} and Q_{ev}, there exist a sequence ω, a lower semicomputable semimeasure P, and a partial computable F such that for all n

$$(Q_{\text{odd}}Q_{\text{ev}})(\omega_{1\ldots 2n}) \leq (3/4)^n P(\omega_{1\ldots 2n})$$

and $F(\omega_{1\ldots 2n}, \omega_{2n+2}) = \omega_{2n+1}$.

▶ Proposition 7. For all lower semicomputable semimeasures Q, there exist $\alpha > \sqrt{1/2}$ and a family of odd and even semimeasures $P_{\text{odd},n}$ and $P_{\text{ev},n}$ uniformly lower-semicomputable in n, such that for all x

$$P_{\text{odd},|x|}(x) P_{\text{ev},|x|}(x) \geq \alpha^{|x|} Q(x)/4 \,. \qquad (3)$$

Proof that Proposition 7 implies Theorem 3. Choose $Q = M$ in Proposition 7 and let for a sufficiently small $c > 0$

$$P_{\text{odd}}(x) = c\left(\frac{1}{1^2} P_{\text{odd},1}(x) + \frac{1}{2^2} P_{\text{odd},2}(x) + \ldots\right).$$

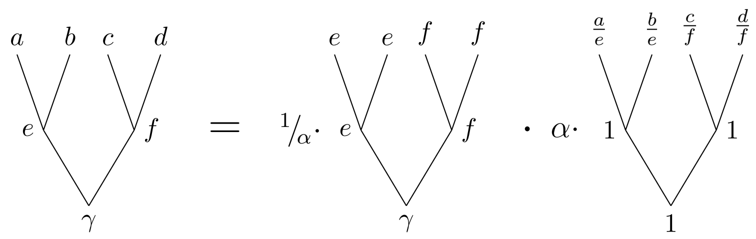

Figure 1 Decomposing semimeasures into odd and even ones.

Note that P_{odd} is a lower semicomputable odd semimeasure and by universality $P_{\text{odd}}(x) \leq O(M_{\text{odd}}(x))$. Hence $-\log M_{\text{odd}}(x) \leq -\log P_{\text{odd},|x|}(x) + O(\log|x|)$. Similar for $P_{\text{ev}}(x)$. By the online coding theorem we obtain up to terms $O(\log|x|)$,

$$(C_{\text{odd}} + C_{\text{ev}})(x) \leq -\log\left(P_{\text{odd},|x|}(x)P_{\text{ev},|x|}(x)\right) \leq -|x|\log\alpha - \log Q(x)\,.$$

Here, $-\log\alpha < 1/2$ and the last term is bounded by $-\log M(x) \leq C(x) + O(\log|x|)$. The $O(\log|x|)$ can be removed for large $|x|$ by choosing $-\log\alpha < \beta < 1/2$. ◀

Proof that Proposition 6 implies Theorem 2. Choosing $Q_{\text{odd}} = M_{\text{odd}}$ and $Q_{\text{ev}} = M_{\text{ev}}$, the first part is immediate by the coding theorem and (2) follows directly from the definition of even complexity. For any x we have

$$C_{\text{odd}}(x) - O(1) \leq C(x) \leq C_{\text{odd}}(x) + C_{\text{ev}}(x) + O(\log|x|)$$

We obtain (1) by applying $C_{\text{ev}}(x) \leq O(1)$. ◀

Proof that Proposition 5 implies Theorem 1. For Theorem 1 we also apply Proposition 5 with $Q_{\text{odd}} = M_{\text{odd}}$ and $Q_{\text{ev}} = M_{\text{ev}}$ to obtain for some $\delta' > 0$

$$(C_{\text{odd}} + C_{\text{ev}})(\omega_{1\ldots2n}) \geq (2 - 2\varepsilon)C(\omega_{1\ldots2n}) + \delta' n\,.$$

Notice that $C_{\text{odd}} \leq C + O(1)$, hence $C_{\text{ev}}(\omega_{1\ldots2n}) \geq (1 - 2\varepsilon)C(\omega_{1\ldots2n}) + \delta' n$; and similar for C_{odd}. Conditions (1) and (2) follow in a similar way as above. ◀

The generalization of Theorem 1 mentioned in section 2 is shown in the full version. We remark that P in these theorems can not be computable, this follows from the subsequent lemma.

▶ **Lemma 8.** *For every computable semimeasure P, there exist computable odd and even online semimeasures P_{odd} and P_{ev} such that $P_{\text{odd}}P_{\text{ev}} = P$.*

Proof. Let ε be the empty string and let $P_{\text{odd}}(\varepsilon) = P(\varepsilon)$ and $P_{\text{ev}}(\varepsilon) = 1$. Suppose that at some node x we have defined $P_{\text{odd}}(x)$ and $P_{\text{ev}}(x)$ such that $P_{\text{odd}}(x)P_{\text{ev}}(x) = P(x)$. Then P_{odd} and P_{ev} are defined on 2-bit extensions of x according to Figure 1 for $\gamma = P(x)$ and $\alpha = P_{\text{ev}}(x)$ [our assumption implies $P_{\text{odd}}(x) = \gamma/\alpha$]. Note that P_{odd} and P_{ev} are indeed computable odd and even semimeasures and that $P_{\text{odd}}P_{\text{ev}} = P$. ◀

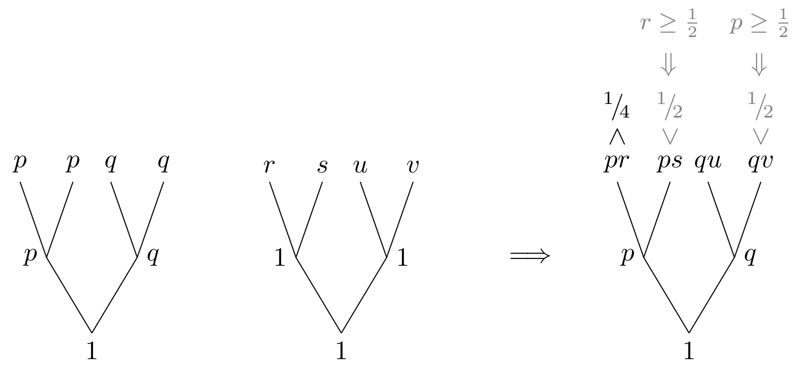

Figure 2 Game for Proposition 6 with $n = 1$.

4 Proofs of lower bounds

We start with Proposition 6, and repeat it for convenience.

▶ Proposition. For all lower semicomputable odd and even online semimeasures Q_{odd} and Q_{ev}, there exist a sequence ω, a lower semicomputable semimeasure P, and a partial computable F such that for all n

$$(Q_{\text{odd}}Q_{\text{ev}})(\omega_{1\ldots 2n}) \leq (3/4)^n P(\omega_{1\ldots 2n})$$

and $F(\omega_{1\ldots 2n}, \omega_{2n+2}) = \omega_{2n+1}$.

To develop some intuition, we first consider a game. The game is played between two players (Alice and Bob) who alternate turns. Alice maintains values for $P(x)$ on 2-bit x. At each round she might pass or increase some values as long as $\sum\{P(x) : |x| = 2\} = 3/4$. Bob maintains lower semicomputable odd and even semimeasures $Q_{\text{odd}}(x)$ and $Q_{\text{ev}}(x)$, see figure 2. Also Bob might pass or increase some values as long as the conditions of the definition of online semimeasure are satisfied, (hence $\max\{p+q, r+s, u+v\} \leq 1$ in figure 2). Alice wins if in the limit $P(x) \geq Q_{\text{odd}}(x)Q_{\text{ev}}(x)$ holds for some x (i.e. if $P(00) \geq pr$ or $P(01) \geq ps$ or $P(10) \geq qu$ or $P(11) \geq qv$).

In this game Alice has a winning strategy. She starts by putting $1/4$ at one leaf and zero at the others, say $P(00) = 1/4$. Then she waits until Bob increases either Q_{odd} or Q_{ev} above $1/2$ at this leaf (thus $Q_{\text{odd}}(0) = Q_{\text{odd}}(00) > 1/2$ or $Q_{\text{ev}}(00) > 1/2$). If none of this happens, Alice wins. Otherwise if $Q_{\text{odd}}(0) > 1/2$, she plays $P(11) = 1/2$ and if $Q_{\text{ev}}(00) > 1/2$, she plays $P(01) = 1/2$. In the first case Alice wins because $Q_{\text{odd}}(1) \leq 1 - Q_{\text{odd}}(0) < 1/2$ and hence $Q_{\text{odd}}(1)Q_{\text{ev}}(11) < 1/2$ and in the second case she wins because $Q_{\text{ev}}(01) \leq 1 - Q_{\text{ev}}(00) < 1/2$ and hence $Q_{\text{odd}}(0)Q_{\text{ev}}(01) < 1/2$. Note that in both cases $\sum\{P(x) : |x| = 2\} = 1/2 + 1/4$, (and otherwise it is $1/4$) and Alice's condition is always satisfied. (Also note that the second bit of x on which Alice wins is 1 if $Q_{\text{odd}}(0) > 1/2$ or $Q_{\text{ev}}(00) > 1/2$. So for lower-semicomputable Q_{odd} and Q_{ev}, we can use this bit to determine which inequality was first realized, and hence to compute the first bit of x. A similar observation will be used to construct F in the proof below.)

To show the proposition, we need to concatenate strategies for the game above to strategies for larger games. For this, it seems that the winning rule needs to be strengthened, and this makes either the winning rule or the winning strategy for the small game complicated. Therefore, in the more concise proof below, we gave a formulation without use of game technique.

Proof. We construct $\omega_{1\ldots 2n}$ together with thresholds o_n, e_n inductively. Let $o_0 = e_0 = 1$. For x of length $2n$, consider the conditions $Q_{\text{odd}}(x0) > o_n/2$ and $Q_{\text{ev}}(x00) > e_n/2$. We fix some algorithm that enumerates Q_{odd} and Q_{ev} from below and after each update tests both conditions. Let O_x be the condition that $Q_{\text{odd}}(x0) > o_n/2$ is true at some update and $Q_{\text{ev}}(x00) > e_n/2$ did not appear at any update strictly before; and let E_x be the condition that $Q_{\text{ev}}(x00) > e_n/2$ is true after some update but $Q_{\text{odd}}(x0) > o_n/2$ is false at the current update (and hence at any update before). Note that O_x and E_x cannot happen both. Let

$$(\omega_{2n+1}\omega_{2n+2}, o_{n+1}, e_{n+1}) = \begin{array}{llll} (11, & o_n/2, & e_n) & \text{if } O_{\omega_{1\ldots 2n}} \text{ happens,} \\ (01, & o_n, & e_n/2) & \text{if } E_{\omega_{1\ldots 2n}} \text{ happens,} \\ (00, & o_n/2, & e_n/2) & \text{otherwise.} \end{array}$$

By induction it follows that $o_n \geq Q_{\text{odd}}(\omega_{1\ldots 2n})$ and $e_n \geq Q_{\text{ev}}(\omega_{1\ldots 2n})$. Indeed, this follows directly for $n = 0$. For $n \geq 1$, consider the case where $O_{\omega_{1\ldots 2n}}$ happens. Thus $\omega_{1\ldots 2n+2} = \omega_{1\ldots 2n+1}1$ and

$$Q_{\text{odd}}(\omega_{1\ldots 2n}1) \leq Q_{\text{odd}}(\omega_{1\ldots 2n}) - Q_{\text{odd}}(\omega_{1\ldots 2n}0) \leq o_n - o_n/2 = o_n/2\,.$$

On the other hand, $Q_{\text{ev}}(\omega_{1\ldots 2n+2}) \leq Q_{\text{ev}}(\omega_{1\ldots 2n}) \leq e_n = e_{n+1}$. The case where $E_{\omega_{1\ldots 2n}}$ happens is similar, and the last one is direct.

It remains to define F and P such that $F(\omega_{1\ldots 2n}, \omega_{2n+2}) = \omega_{2n+1}$ and

$$P(\omega_{1\ldots 2n}) = (4/3)^n o_n e_n\,.$$

Note that $\omega_{2n+2} = 1$ iff $O_{\omega_{1\ldots 2n}}$ or $E_{\omega_{1\ldots 2n}}$ happens, and knowing that one of the events happens, we can decide which one and therefore also ω_{2n+1}. Hence, given $\omega_{1\ldots 2n}$ and ω_{2n+2} we can compute ω_{2n+1} and this procedure defines the partial computable function F.

To define P, observe that ω can be approximated from below: start with $\omega = 00\ldots$, each time $O_{\omega_{1\ldots 2n}}$ (respectively $E_{\omega_{1\ldots 2n}}$) happens, change $\omega_{2n}\omega_{2n+1}$ from 00 to 01 (respectively to 11), let all subsequent bits be zero, and repeat the process. Hence, for all n and $2n$-bit x at most one pair (o_n, e_n) is defined which we denote as (o_x, e_x). Let $P(x)$ be zero unless (o_x, e_x) is defined in which case

$$P(x) = (4/3)^{|x|/2} o_x e_x\,.$$

Note that P is lower semicomputable and the equation above is satisfied. Also, P is a semimeasure: $P(\varepsilon) = (4/3)^0 \cdot 1 \cdot 1 = 1$, and in all three cases we have $\sum\{o_{xbb'}e_{xbb'} : b, b' \in \{0,1\}\} \leq 3o_x e_x/4$ hence, $\sum\{P(xbb') : b, b' \in \{0,1\}\} \leq P(x)$. ◀

The proof of Proposition 5 follows the same structure.

▶ Proposition. For all $\varepsilon > 0$ and lower semicomputable odd and even online semimeasures Q_{odd} and Q_{ev}, there exist δ, a sequence ω, a lower semicomputable semimeasure P, and a partial computable F such that for all n

$$(Q_{\text{odd}}Q_{\text{ev}})(\omega_{1\ldots n}) \leq (1-\delta)^n P(\omega_{1\ldots n})^{2-2\varepsilon}$$

and $F(\omega_{1\ldots 2n}, \omega_{2n+2}) = \omega_{2n+1}$.

Proof. We first consider the following variant for the game above on strings of length two. Alice should satisfy the weaker condition $\sum\{P(x) : |x| = 2\} \leq 1 - \delta$, where $\delta \ll \varepsilon$ will be determined later. She wins if

$$(P_{\text{odd}}P_{\text{ev}})(x) \leq (P(x))^{2-2\varepsilon}$$

for some x. The idea of the winning strategy is to start with a very small value somewhere, say $P(00) = \delta$. If $\varepsilon = 0$ then Bob could reply with $Q_{\text{odd}}(0) = Q_{\text{ev}}(00) = \delta$, (in fact he could win by always choosing $Q_{\text{odd}}(x) = Q_{\text{ev}}(x) = P(x)$). For $\varepsilon > 0$ and $\delta \ll \varepsilon$ one of the online semimeasures should exceed $\delta^{1-\varepsilon} = k\delta$ for $k = \delta^{-\varepsilon}$. k can be arbitrarily large if $\delta \ll \varepsilon$ is chosen sufficiently small. At his next move, (as before), Alice puts all his remaining measure, i.e. $1 - 2\delta$ in a leaf that does not belong to a branch where the corresponding online semimeasure is large. Note that $1 - 2\delta$ is close to 1 and taking a power $2 \geq 2 - 2\varepsilon$ we see that Bob needs at least $1 - 4\delta$ in each online semimeasure, but he already used $k\delta$ in one of them.

More precisely, the winning strategy for Alice is to set $P(00) = \delta$ and wait until $Q_{\text{odd}}(0) > \delta^{1-\varepsilon}$ or $Q_{\text{ev}}(00) > \delta^{1-\varepsilon}$. If these conditions are never satisfied, then Alice wins on $x = 00$. Suppose at some moment Alice observes that the first condition holds, then she plays $P(11) = 1 - 2\delta$, in the other case she plays $P(01) = 1 - 2\delta$. Afterwards she does not play anymore. Note that $\sum\{P(x) : |x| = 2\} \leq 1 - \delta$. We show that Alice wins. Assume that $Q_{\text{odd}}(0) > \delta^{1-\varepsilon}$ (the other case is similar). We know that $Q_{\text{ev}}(11) \leq 1$ hence if Alice does not win, this implies $Q_{\text{odd}}(1) > (1 - 2\delta)^{2-2\varepsilon}$. This is lower bounded by $(1 - 2\delta)^2 \geq 1 - 4\delta$. We choose $\delta = 2^{-2/\varepsilon}$. This implies

$$\delta^{1-\varepsilon} = 2^{-(2/\varepsilon)(1-\varepsilon)} = 2^{-2/\varepsilon+2} = 4\delta.$$

Hence $Q_{\text{odd}}(0) + Q_{\text{odd}}(1) > 4\delta + (1 - 4\delta) = 1$ and Bob would violate his restrictions. Therefore Alice wins. For later use notice that in the first case our argument implies $Q_{\text{odd}}(1) \leq (1 - 2\delta)^{2-2\varepsilon}$.

In a similar way as before we adapt Alice's strategy to an inductive construction of ω and P: let O_x and E_x be defined as before using conditions $Q_{\text{odd}}(x0) > o_n\delta^{1-\varepsilon}$ and $Q_{\text{ev}}(x00) > e_n\delta^{1-\varepsilon}$. Let $\beta = (1 - 2\delta)^{2-2\varepsilon}$ and let ω, o_n and e_n be given by

$$(\omega_{2n+1}\omega_{2n+2}, o_{n+1}, e_{n+1}) = \begin{array}{llll} (11, & o_n\beta, & e_n) & \text{if } O_{\omega_{1\ldots 2n}} \text{ happens,} \\ (01, & o_n, & e_n\beta) & \text{if } E_{\omega_{1\ldots 2n}} \text{ happens,} \\ (00, & o_n\delta^{1-\varepsilon}, & e_n\delta^{1-\varepsilon}) & \text{otherwise.} \end{array}$$

This implies $o_n \geq Q_{\text{odd}}(\omega_{1\ldots 2n})$ and $e_n \geq Q_{\text{ev}}(\omega_{1\ldots 2n})$. F is defined and shown to satisfy the condition in exactly the same way. It remains to construct P such that

$$(1 - \delta)^n P(\omega_{1\ldots 2n}) = (o_n e_n)^{1/(2-2\varepsilon)} \;,$$

(the proposition follows after rescaling δ). In a similar way as before o_x and e_x are defined and let

$$P(x) = (1 - \delta)^{-|x|/2}(o_x e_x)^{1/(2-2\varepsilon)} .$$

To show that P is indeed a semimeasure observe that $\sum\{P(xbb') : b, b' \in \{0, 1\}\}$

$$\begin{aligned} &= (1 - \delta)^{-|x|/2-1} \sum\{(o_{xbb'}e_{xbb'})^{1/(2-2\varepsilon)} : b, b' \in \{0, 1\}\} \\ &\leq (1 - \delta)^{-|x|/2-1} \left(\beta^{1/(2-2\varepsilon)} + \delta\right) (o_x e_x)^{1/(2-2\varepsilon)} \;, \end{aligned}$$

and because $\beta^{1/(2-2\varepsilon)} = 1 - 2\delta$ this equals

$$= (1 - \delta)^{-|x|/2} \,(o_x e_x)^{1/(2-2\varepsilon)} = P(x) .$$ ◀

Acknowledgements. The author is grateful to Alexander Shen, Nikolay Vereshchagin, Andrei Romashchenko, Mikhail Dektyarev, Ilya Mezhirov and Emmanuel Jeandel for extensive discussion and many useful suggestions. I also thank Ilya Mezhirov for implementing clever code to study some games. Especially thanks to Alexander Shen for encouragement after presenting earlier results and for arranging funding by grant NAFIT ANR-08-EMER-008-01. The author is also grateful to Mathieu Hoyrup who arranged a grant under which the work was finalized.

References

1 B. Bauwens and A. Shen. An additivity theorem for plain Kolmogorov complexity. *Theory Computing Systems*, 52(2):297–302, 2013.

2 Bruno Bauwens. *Computability in statistical hypotheses testing, and characterizations of independence and directed influences in time series using Kolmogorov complexity.* PhD thesis, Ghent University, May 2010.

3 A. Chernov, A. Shen, N. Vereshchagin, and V.Vovk. On-line probability, complexity and randomness. In *ALT'08: Proceedings of the 19th international conference on Algorithmic Learning Theory*, pages 138–153, Berlin, Heidelberg, 2008. Springer-Verlag.

4 U. Feldmann and J. Bhattacharya. Predictability improvement as an asymmetrical measure of interdependence in bivariate time series. *International Journal of Bifurcation and Chaos*, 14(2):505–514, 2004.

5 P. Gács. Lecture notes on descriptional complexity and randomness. `http://www.cs.bu.edu/faculty/gacs/papers/ait-notes.pdf`, 1988–2011.

6 C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

7 G. John. Inference and causality in economic time series models. In Z. Griliches and M. D. Intriligator, editors, *Handbook of Econometrics*, volume 2, chapter 19, pages 1101–1144. Elsevier, 1984.

8 M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer-Verlag, New York, 2008.

9 An. A. Muchnik. Algorithmic randomness and splitting of supermartingales. *Problems of Information Transmission*, 45(1):54–64, March 2009.

10 M. Palus and A. Stefanovska. Direction of coupling from phases of interacting oscillators: an information theoretic approach. *Physical Review E, Rapid Communications*, 67:055201(R), 2003.

11 A. Papana, C. Kyrtsou, D. Kugiumtzis, and C. Diks. Simulation study of direct causality measures in multivariate time series. *Entropy*, 15(7):2635–2661, 2013.

12 J. Pearl. *Causality: Models, Reasoning and Inference.* Cambridge University Press, 2000.

13 F.A. Razak. *Mutual information based measures on complex interdependent networks of neuro data sets.* PhD thesis, Imperial College London, March 2013.

14 M. G. Rosenblum and A. S. Pikovsky. Detecting direction of coupling in interacting oscillators. *Phys. Rev. E*, 64(4):045202, Sep 2001.

15 T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461–464, Jul 2000.

16 V. A. Uspensky, N. K.Vereshchagin, and A. Shen. Kolmogorov complexity and algorithmic randomness. To appear.

17 V. A. Uspensky and A. Shen. Relations between varieties of Kolmogorov complexities. *Theory of Computing Systems*, 29(3):271–292, 1996.

18 M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistritz, R. Bauer, J. Timmer, and H. Witte. Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. *Signal Processing*, 85:2137–2160, 2005.

19 A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6:156):83–124, 1970.