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

Bruno Bauwens

Université de Lorraine, LORIA
615, Rue du Jardin Botanique, France
Brbauwens at gmail dot 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 $x_1x_2 \ldots x_n$ 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 $x_2x_1x_4x_3 \ldots$, decreases the sum of odd and even complexity to $C(x)$. Our result is related to the problem of inference of causality in timeseries.

1998 ACM Subject Classification E.4

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

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

For the interactive case, we need the online variant of Kolmogorov complexity introduced in [3]. Let \( C_{ev}(x) \) 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 \( C_{odd}(x) \). 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 2, we show that there exist infinitely many bitstrings \( x \), such that both \( C_{ev}(x) \) and \( C_{odd}(x) \) are almost as big as \( C(x) \), in agreement with our example. In Theorem 2, we show that there exists \( c > 0 \) such that \( (C_{ev} + C_{odd} - C)(x) \geq c|x| \), i.e. the online asymmetry of information can be large compared to the length of \( x \). Finally, we raise the question how large \( (C_{ev} + C_{odd} - C)(x) \) 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 \( (C_{ev} + C_{odd} - C)(x) \leq (1/2 - c)|x| \) 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 \( C_{odd} + C_{ev} \) 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.

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. For time delays \( k \) and \( l > k \) the information transfer from \( X \) to \( Y \) is given by

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

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.
In our example, coin tosses (the conditional version) of symmetry of information. In the offline case, a similar observation if this term is dependent on \( n \), the sum is taken. This quantification of causality coincides with Granger causality \( \Omega \) if all involved conditional distributions are Gaussian.

We show that for all \( x \in \mathcal{X} \) and \( y \in \mathcal{Y} \) the information transfers from \( \mathcal{X} \) to \( \mathcal{Y} \) and \( \mathcal{Y} \) to \( \mathcal{X} \) can be different. On the other hand, if \( k = 0 \) 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: 

\[
IT(x \to y) = C(y) - C_{ev}(x_1y_1 \ldots x_ny_n).
\]

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 \( x_1 \) 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 \( \Omega \) that is related to our observations. Consider the example from \( \Omega \): in a tournament (say chess), a coin toss decides which player starts the next game. Consider the sequence \( b_1, w_1, b_2, w_2, \ldots \) 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 \( \omega \) 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 \( b_1w_1b_2w_2 \ldots \) 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, then lower bounds are proven, and finally (in Appendix section \( \Omega \)) the upper bound is proven. In the next appendix we generalize Theorem \( \Omega \) for online computation.

---

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 \( K(K(x)) - K(x)|x| \leq O(1) \) while \( K(x) - K(x|x|) \geq \log n - O(\log \log n) \). Moreover, this small error was exploited in an earlier and more involved proof of Theorem \( \Omega \).
with more machines and present a more involved proof of the main result with slightly better parameters. In the last appendix we generalize symmetry of information to a chain rule for online Kolmogorov complexity.

## 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] for details. If \( y \) is the empty string, we write \( C_U(x) \) in stead of \( C_U(x|y) \), 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

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

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 \( C(x|y) - O(1) \leq C_{ev}(y_1x_1 \ldots y_nx_n) \leq C(x) + O(1) \) for \( n \)-bit \( x \) and \( y \). Let \( C_{ev}(w|v) \) be the conditional variant. The chain rule for the concatenation \( vw \) of strings \( v \) and \( w \) holds: \( C_{ev}(vw) = C_{ev}(v) + C_{ev}(w|v) + O(\log(|v|)) \), see Appendix D.

In a similar way \( C_{odd}(x) \) is defined. A direct lower and upper bound for \( C_{odd} + C_{ev} \) are

\[
C(z) - O(\log |z|) \leq (C_{odd} + C_{ev})(z) \leq 2C(z) + O(1).
\]

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 \( C_{odd} + C_{ev} \) can change significantly after a simple permutation of the bits.

\textbf{Theorem 1.} For every \( \varepsilon > 0 \) there exist \( \delta > 0 \) and a sequence \( \omega \) such that for large \( n \)

\[
\frac{C_{odd}(\omega_1 \ldots \omega_n)}{C_{ev}(\omega_1 \ldots \omega_n)} \geq (1 - \varepsilon)C(\omega_1 \ldots \omega_n) + \delta n.
\]

Moreover, for all even \( n \)

\[
\begin{align*}
C_{odd}(\omega_2\omega_1 \ldots \omega_n\omega_{n-1}) &= C(\omega_1 \ldots \omega_n) + O(\log n) \quad (1) \\
C_{ev}(\omega_2\omega_1 \ldots \omega_n\omega_{n-1}) &\leq O(1). \\
\end{align*}
\]

The first part implies

\[
\limsup_{|z| \to \infty} \frac{C_{odd}(x) + C_{ev}(x)}{C(x)} \geq 2,
\]

The \( O(\log |x|) \) term could be decreased to \( O(1) \) if we compared online complexity with decision complexity [12] 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 $IT(x \to y) = C(y) - C_{\text{ev}}(x_1 y_1 \ldots x_n y_n)$ for $x, y, n$ such that $n = |x| = |y|$. Let $x = \omega_1 \omega_3 \ldots \omega_{2n-1}$ and $y = \omega_2 \omega_4 \ldots \omega_{2n}$, Theorem 3 implies

\[
IT(x \to y) \leq \varepsilon C(x, y) + O(1) \\
IT(y \to x) = C(x, y) + O(1),
\]

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

Theorem 2 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 $i \mod k$ be given by

\[
C_i \mod k(x) = \min \{|p| : U(p, x_1 \ldots x_{j-1}) = x_j \text{ for all } j = i, i + k, \ldots, \leq |x|\}.
\]

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

\[
C_i \mod k(\omega_1 \ldots \omega_n) \geq (1 - \varepsilon)C(\omega_1 \ldots \omega_n) + \delta n
\]

Moreover, for $\bar{\omega} = \omega_1 \omega_1 \ldots \omega_{k-1} \omega_2 \omega_{k+1} \ldots \omega_{2k-1} \ldots$ for all $n$, and $i = 2 \ldots k$:

\[
C_1 \mod k(\bar{\omega}_1 \ldots \bar{\omega}_n) = C(\omega_1 \ldots \omega_n) + O(\log n)
\]

In Theorem 2 the difference between $C$ and $C_{\text{odd}} + C_{\text{ev}}$ is linear in the length of the prefix of $\omega$. 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

\[
(C_{\text{odd}} + C_{\text{ev}})(x) - C(x) = (C_{\text{odd}}(x) - C(x)) + C_{\text{ev}}(x) \leq O(1) + |x|/2 + O(1).
\]

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

**Theorem 2.** There exist a sequence $\omega$ such that for all $n$

\[
(C_{\text{odd}} + C_{\text{ev}})(\omega_1 \ldots \omega_n) \geq n(\log \frac{2}{3})/2 + C(\omega_1 \ldots \omega_n) - O(\log n).
\]

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

In the appendix we show how the factor $(\log \frac{2}{3})/2$ can further be improved to $(\log \frac{2}{3})/2 \approx 0.292$ at the cost of weakening (1) and (2). 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 $x$

\[
(C_{\text{ev}} + C_{\text{odd}} - C)(x) \leq \beta|x|.
\]

In summary,\footnote{For the first we use $C(y) \leq C(\omega_1 \ldots \omega_n) = C(x, y)$ up to $O(1)$ terms. For the second $C(x, y) \geq C(x) \geq C_{\text{ev}}(y_1 x_1 \ldots x_n y_n) = C(y, x)$, thus $C(x) = C(x, y)$, while $C_{\text{ev}}(y_1 x_1 \ldots y_n x_n) \leq O(1)$. Also, note that $C(\omega_1 \ldots \omega_n)$ must exceed $\delta n$ because it exceeds $C_{\text{odd}}(\omega_1 \ldots \omega_n) \geq \delta n$, all up to $O(1)$ terms.} $\frac{1}{2} \log \frac{3}{2} \leq \limsup \frac{(C_{\text{ev}} + C_{\text{odd}} - C)(x)}{|x|} < \frac{1}{2}$, but the precise value of the lim sup is unknown.
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 $P_1, P_2, \ldots$ 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 $M_{\text{ev}}$ (resp. $M_{\text{odd}}$) satisfy

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

Let $\omega_{k\ldots 1} = \omega_k \ldots \omega_1$. Theorems 1 and 3 follow from

Proposition 5. For all $\varepsilon > 0$ and lower semicomputable odd and even online semimeasures $Q_{\text{odd}}$ and $Q_{\text{ev}}$, there exist $\delta$, a sequence $\omega$, 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 $Q_{\text{odd}}$ and $Q_{\text{ev}}$, there exist a sequence $\omega$, 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 $x$

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

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

$$P_{\text{odd}}(x) = c \left( \frac{1}{12} P_{\text{odd,1}}(x) + \frac{1}{22} P_{\text{odd,2}}(x) + \ldots \right).$$
We obtain (1) by applying Proposition 6, and repeat it for convenience. The generalization of Theorem 1 mentioned in section 2 is shown in the appendix. We remark with $Q = \alpha$. Let $P$ be some node online semimeasures and that $\alpha = P(\gamma)$. Then $P(\gamma)$, the online coding theorem we obtain up to terms $O(\log |x|)$,

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

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 $\delta' > 0$

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

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

The generalization of Theorem 1 mentioned in section 2 is shown in the appendix. 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 $P_{\text{odd}}$ and $P_{\text{ev}}$ such that $P_{\text{odd}}P_{\text{ev}} = P$.

**Proof.** Let $\varepsilon$ 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 $P_{\text{odd}}$ and $P_{\text{ev}}$ 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 $P_{\text{odd}}$ and $P_{\text{ev}}$ are indeed computable odd and even semimeasures and that $P_{\text{odd}}P_{\text{ev}} = P$.

4 Proofs of lower bounds

We start with Proposition 6 and repeat it for convenience.
Asymmetry of online Kolmogorov complexity

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

$$(Q_{\text{odd}}Q_{\text{ev}})(\omega_{1\ldots2n}) \leq (3/4)^n P(\omega_{1\ldots2n})$$

and $F(\omega_{1\ldots2n},\omega_{2n+2}) = \omega_{2n+1}$.

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 $Q_{\text{odd}}$ and $Q_{\text{ev}}$, 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 $Q_{\text{odd}}$ or $Q_{\text{ev}}$ 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 $Q_{\text{odd}}$ and $Q_{\text{ev}}$, 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 $\omega_{1\ldots2n}$ together with thresholds $\alpha_n, \epsilon_n$ inductively. Let $\alpha_0 = \epsilon_0 = 1$. For $x$ of length $2n$, consider the conditions $Q_{\text{odd}}(x0) > \alpha_n/2$ and $Q_{\text{ev}}(x00) > \epsilon_n/2$. We
fix some algorithm that enumerates $Q_{\text{odd}}$ and $Q_{\text{ev}}$ from below and after each update tests both conditions. Let $O_x$ be the condition that $Q_{\text{odd}}(x0) > a_n/2$ 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 $Q_{\text{odd}}(x0) > a_n/2$ is false at the current update (and hence at any update before). Note that $O_x$ and $E_x$ cannot happen both. Let

\[
(\omega_{2n+1}\omega_{2n+2}, o_{n+1}, e_{n+1}) =
\begin{cases}
(11, \ a_n/2, \ e_n) & \text{if } Q_{\omega_{1\ldots 2n}} \text{ happens}, \\
(01, \ a_n, \ e_n/2) & \text{if } E_{\omega_{1\ldots 2n}} \text{ happens}, \\
(00, \ a_n/2, \ e_n/2) & \text{otherwise}.
\end{cases}
\]

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 $\omega_{1\ldots 2n+2} = \omega_{1\ldots 2n+1}$ and

\[
Q_{\text{odd}}(\omega_{1\ldots 2n+1}) \leq Q_{\text{odd}}(\omega_{1\ldots 2n}) - Q_{\text{odd}}(\omega_{1\ldots 2n}0) \leq o_n - a_n/2 = o_n/2.
\]

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

To define $P$, observe that $\omega$ 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: $P(\epsilon) = (4/3)^0 \cdot 1 = 1$, and in all three cases we have $\sum (o_{b'b'} e_{b'b'} : b, b' \in \{0, 1\}) \leq 3 o_x e_x / 4$ hence, $\sum P(xb'b') : b, b' \in \{0, 1\} \leq P(x)$.}

The proof of Proposition 5 follows the same structure.

**Proposition.** For all $\varepsilon > 0$ and lower semicomputable odd and even online semimeasures $Q_{\text{odd}}$ and $Q_{\text{ev}}$, there exist $\delta$, a sequence $\omega$, 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

\[
(P_{\text{odd}} P_{\text{ev}})(x) \leq (P(x))^{2-2\varepsilon}
\]
Asymmetry of online Kolmogorov complexity

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 $\delta < \varepsilon$ one of the online semimeasures should exceed $\delta^{1-\varepsilon} = k\delta$ for $k = \delta^{-\varepsilon}$. $k$ can be arbitrarily large if $\delta < \varepsilon$ 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

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

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 $\omega$ 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 $\omega, o_n$ and $e_n$ be given by

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

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

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

(the proposition follows after rescaling $\delta$). 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 $\sum\{P(xbb') : b, b' \in \{0, 1\}\}$

$$(1 - \delta)^{-|x|/2 - 1} \sum\{(o_x b \delta e_x b')^{1/(2 - 2\varepsilon)} : b, b' \in \{0, 1\}\}$$

$$\leq (1 - \delta)^{-|x|/2 - 1} \left(\beta^{1/(2 - 2\varepsilon)} + \delta\right)(o_x e_x)^{1/(2 - 2\varepsilon)},$$

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

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

The author is grateful to Alexander Shen, Nikolay Vereshchagin, Andrei Romashchenko, Mikhail Dehtyarev, 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. Computability in statistical hypotheses testing, and characterizations of independence and directed influences in time series using Kolmogorov complexity. PhD thesis, UGent, May 2010.
2. B. Bauwens and A. Shen. An additivity theorem for plain Kolmogorov complexity. Theory Computing Systems, 52(2):297–302, 2013.
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 descriptive complexity and randomness. [http://www.cs.bu.edu/faculty/gacs/papers/alt-notes.pdf](http://www.cs.bu.edu/faculty/gacs/papers/alt-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. 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. Kyrtou, 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.
Asymmetry of online Kolmogorov complexity

12

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.

A Proof of the upper bound: Theorem 3

It remains to prove Proposition 7, which we repeat.

**Proposition.** 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 $x$

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

Consider the game defined before the proof of Proposition 6 on 2-bit $x$, where we replace $3/4$ by any real number $\alpha^2$. Thus, Alice enumerates non-decreasing values $P(x)$ such that $P(x) \leq \alpha^2$ and Bob enumerates $P_{\text{odd}}(x)$ and $P_{\text{ev}}(x)$ for all two-bit $x$. Bob wins if $P_{\text{odd}}(x)P_{\text{ev}}(x) > P(x)$ for all $x$. It can be shown that Bob has a winning strategy in this game if $\alpha^2 < 2/3$, and his strategy is uniformly computable in $\alpha$. (Otherwise Alice wins, see appendix C). Proposition 7 would follow for all $\alpha^2 < 2/3$ if induction schemes for these strategies exist. Unfortunately, we do not know such schemes. Therefore, a game is formulated with a stronger winning condition for Bob that allows to concatenate winning strategies. We show a winning strategy for some $\alpha^2$ (significantly below 2/3 and) above 1/2.

**Proof.** Let $\varepsilon > 0$ be small to be determined later. For every $n$ and every lower semicomputable semimeasure $P_n$ defined on all $n$-bit $x$ we construct lower semicomputable $P_{\text{odd}}$ and $P_{\text{ev}}$ such that $(P_{\text{odd}}P_{\text{ev}})(x) \geq (1/\sqrt{2} + \varepsilon/2)^nP_n(x)/4$ (on all $n$-bit $x$). Moreover, our construction is uniform in $n$ and $P_n$ and this implies Proposition 7.

We will represent $P_n$ by a $2^n$-dimensional real vector $u$ such that $u_1, u_2, \ldots, u_{2^n}$ equal the values of $\sqrt{P_n(x)}$ on all $n$-bit $x$ in lexicographic order. Note that the definition of semimeasure implies $\|u\|_\infty \leq 1$.

We construct online semimeasures $P_{\text{odd}}$ and $P_{\text{ev}}$ from $2^n$-dimensional vectors $o$ and $p$ by defining $P_{\text{odd}}$ and $P_{\text{ev}}$ to be the smallest odd and even online semimeasures whose values on all $n$-bit $x$ in lexicographic order do not exceed $o_1, \ldots, o_{2^n}$ and $p_1, \ldots, p_{2^n}$. Note that such $P_{\text{odd}}$ and $P_{\text{ev}}$ can be computed from $o$ and $p$ because for fixed $n$ only finitely many binary max and sum operations appear in the computation.

To see whether some $2^n$-dimensional vectors $o$ and $p$ indeed define such semimeasures, consider all large enough functions $P_n$ and $P_{\text{odd}}$ satisfying conditions $i$ and $ii$ in the definition of online semimeasures. Let us derive these minimal values for $P_{\text{odd}}(\varepsilon)$ and $P_{\text{ev}}(\varepsilon)$ (which should be at most one). For the case $n = 1$, $o$ and $p$ are 2-dimensional and the minimal root values are $\|o\|_\infty$ and $\|p\|_\infty$. For $n > 1$, note that in the tree representation of $P_{\text{odd}}$ and $P_{\text{ev}}$, a value of a node is lower bounded by either the max or the sum of its child values (see figure 3). Hence, we define the norms $\|\cdot\|_\infty$ and $\|\cdot\|_1$ inductively: for a 1-dimensional vector $u$ let $\|u\|_\infty = \|u\|_\infty = u_1$. For an even dimensional vector $u$ let $u_+$ and $u_-$ denote the first and last half of indices of $u$. For a $2n+1$-dimensional vector $u$ let

$$\|u\|_\infty = \max\{\|u_+\|_\infty, \|u_-\|_\infty\}$$

$$\|u\|_1 = \min\{\|u_+\|_1, \|u_-\|_1\}.$$
We do this in an effective way and guarantee that if some coordinates of \( u \) are non-decreasingly updated, then also \( o \) and \( p \) have non-decreasing updates. These updates might depend on the history of updates for \( u \).

Note that \( \| \cdot \|_\infty \) and \( \| \cdot \|_{1,\infty} \) indeed define norms. We construct functions \( o \) and \( p \) such that

1. \( \|o(u)\|_{\infty} \leq \|u\|_2 \),
2. \( \|p(u)\|_{1,\infty} \leq \|u\|_2 \),
3. \( o(u)p(u) \geq (1/\sqrt{2} + \varepsilon/2)^n u^2 \),

(vectors are multiplied point wise). We do this in an effective way and guarantee that some coordinates of \( u \) are non-decreasingly updated, then also \( o \) and \( p \) have non-decreasing updates.

These updates might depend on the history of updates for \( u \). By the above discussion, the functions \( o \) and \( p \) define the requested lower semicomputable online semimeasures.

It remains to construct \( o \) and \( p \). For 1-dimensional \( u \) (i.e. \( n = 0 \)) we choose \( o(u) = p(u) = u \). Clearly the conditions are satisfied. We explain the induction step. Suppose the case \( n = 1 \) is solved. In fact, for small \( \varepsilon \) our solution will be approximately \( o(u) = (1 + a_n \varepsilon) u \) and \( p(u) = (1 + b_n \varepsilon) u/\sqrt{2} \) for some (piecewise constant) rational numbers \( a_n \) and \( b_n \) such that \( a_n + b_n \geq 1 \). Let \( \circ \) denote concatenation; thus \( u = u \circ u \). For the (approximate) induction step, first evaluate \( o(u), p(u), o(u \circ u), p(u \circ u) \) and \( v_u = [\|u\|_2, \|u \circ u\|_2] \). Then combine it using the factors from the \( n = 1 \) case: \( o(u) = (1 + a_n \varepsilon)(p(u \circ u)) \) and similar for \( p(u) \).

More precisely, we construct functions \( \bar{o}(v, u) \) and \( \bar{p}(v, u) \) defined for 2-dimensional \( v \) and real vectors \( u \) of any dimension. (The functions are obtain from the \( n = 1 \) step.) \( o \) and \( p \) for \( u \) of dimension \( 2^n \geq 2 \) are inductively defined by

\[
\begin{align*}
o(u) &= \bar{o}(v_u, p(u \circ u)) \\
p(u) &= \bar{p}(v_u, o(u \circ u))
\end{align*}
\]

The computation of \( p(u) \) is illustrated in figure 3 for \( n = 3 \). The functions \( \bar{o} \) and \( \bar{p} \) satisfy the following properties:

- if \( \|u\|_{\infty} \leq v_0 \) and \( \|u \circ u\|_{\infty} \leq v_1 \), then \( \|\bar{p}(v, u)\|_{1,\infty} \leq \|v\|_2 \),
- if \( \|u\|_{\infty} \leq v_0 \) and \( \|u \circ u\|_{\infty} \leq v_1 \), then \( \|\bar{o}(v, u)\|_{\infty} \leq \|v\|_2 \),
- \( (\bar{o}\bar{p})(v, u) \geq (1/\sqrt{2} + \varepsilon/2) u^2 \),
Asymmetry of online Kolmogorov complexity

Remind that $v_u = [\|u_+\|_2, \|u_-\|_2]$, thus $\|v_u\|_2 = \|u\|_2$. By induction, $o$ and $p$ satisfy properties 1 and 2, and property 3 follows directly.

We now construct $o(u)$ and $p(u)$ for 2-dimensional $u$ such that conditions [1]-[3] are satisfied; afterwards we finish the proof by generalizing the construction to obtain $\overline{o}$ and $\overline{p}$ satisfying the conditions of the induction step.

Let us try the following solution: $o(u) = u$ and $p(u) = \alpha u$ for some $\alpha > 1/\sqrt{2}$. Clearly, [1] and [3] are satisfied (for some $\epsilon > 0$). Consider the second condition: $u_0 + u_1 \leq \sqrt{u_0^2 + u_1^2}$ (2-dimensional vectors are labeled by 0 and 1, rather than 1 and 2). For $u_0 = u_1$ this can only hold if $\alpha \leq 1/\sqrt{2}$, violating the assumption. On the other hand, in this special case [3] is satisfied within a large margin. See figure 5 for an illustration of the conditions. Hence let us try the solution $o(u) = \alpha u$ and $p(u) = u/\sqrt{2}$ for some $\alpha > 1$. Now [2] and [3] are satisfied. Choose $u_0 = 0$ and $u_1 = 1$, the first condition implies $\alpha \leq 1$, again violating the assumption. On the other hand, [2] is satisfied within a large margin.

These observations suggest to construct $o$ and $p$ differently according to the region where $u$ is located: if $u_0 \approx u_1$ then $p$ should be relatively small while $o$ is large and if $u_0/u_1$ or $u_1/u_0$ is small, then $p$ should be relatively large while $o$ is small. In the balanced region (i.e. where $u_0 \approx u_1$), $p$ grows slowly at updates of $u$, so that at the critical line $u_0 = u_1$ the excess from the unbalanced region is compensated and $p$ is relatively small. In order to satisfy [3], $o$ must grow faster. In the unbalanced region, $o$ grows slower so that if $u_0/u_1$ or $u_1/u_0$ become small, the excess from the balanced stage is compensated. To satisfy [3] we need that $p$ grows fast. Despite the simple idea, the precise calculations seem laborious.

Let $u = [a, b]$. $u$ is balanced iff $a/\sqrt{20\epsilon} \leq b \leq a/\sqrt{20\epsilon}$, otherwise $u$ is unbalanced. At each stage let the values of $o, p, u, \ldots$ (we drop arguments $u$) after the previous stage be denoted as $o_o, p_o, u_o, \ldots$, where a subscript $o$ abbreviates “old”. For unbalanced $u$ let

\[
\begin{align*}
    p &= (1 + 2\epsilon)u/\sqrt{2} \\
    o &= \max \{o_o, (1 - \epsilon)u\},
\end{align*}
\]

and if $u$ is balanced let

\[
\begin{align*}
    p &= \max \{p_o, (1 - 4\epsilon)u/\sqrt{2}\} \\
    o &= (1 + 5\epsilon)u.
\end{align*}
\]

In these definitions the max functions guarantee that $o$ and $p$ are non-decreasing in those regions where they grow slower. For small $\epsilon$ condition [3] is always satisfied. We need to check [1] and [2] in the unbalanced an balanced stage.
Assume the case holds for \( \beta = 0.491 \).

1) Balanced \( \|o\|_\infty \leq \|u\|_2 \), i.e. \( o_0 \leq \|u\|_2 \) and \( o_1 \leq \|u\|_2 \). We show the first one (the other is similar):

\[
(1 + 5\varepsilon)a \leq \sqrt{a^2 + b^2},
\]

i.e. \( 1 + 5\varepsilon \leq \sqrt{1 + (b/a)^2} \). Because \( b^2 \geq 20\varepsilon a^2 \) the right hand side is at least \( 1 + 10\varepsilon - O(\varepsilon^2) \), and this exceeds the left hand side for small \( \varepsilon \).

2) Unbalanced \( \|p\|_1 \leq \|u\|_2 \):

\[
(1 + 2\varepsilon)(a + b)/\sqrt{2} \leq \sqrt{a^2 + b^2}.
\]

Rearranging:

\[
(1 + 2\varepsilon)/\sqrt{2} \leq \frac{\sqrt{a^2 + b^2}}{a + b}.
\]

Assume the case \( b/a \leq 20\varepsilon \), the other case is similar. Dropping \( b^2 \) in the right-hand, we have \( 1/(1 + b/a) \) which is bounded by \( 1 - \sqrt{20\varepsilon} + O(\varepsilon) \). Hence it exceeds the left hand for small \( \varepsilon \).

3) Unbalanced \( \|o\|_\infty \leq \|u\|_2 \), i.e. \( o_0 \leq \|u\|_2 \) and \( o_1 \leq \|u\|_2 \). Only the first is shown, the second is similar. Suppose no balanced stage has happened, then \( o \leq (1 - \varepsilon)u \). Otherwise, let \( u_o \) be the value of \( u \) at the last balanced stage. In our upper bound for \( o \) there might be missing at most an excess \( (1 + 5\varepsilon)u_o - (1 - \varepsilon)u_o = 6\varepsilon u_o \). Thus, the equation becomes

\[
(1 - \varepsilon)a + 6\varepsilon a_o \leq \sqrt{a^2 + b^2}.
\]

Note that \( b \geq b_o \geq a_o\sqrt{20\varepsilon} \), hence for \( x = a_o/a \leq 1 \) this follows from

\[
1 - \varepsilon + 6\varepsilon x \leq \sqrt{1 + 20\varepsilon x^2} = 1 + 10\varepsilon x^2 + O(\varepsilon^2)
\]

i.e. \( 0 \leq 10\varepsilon^2 - 6\varepsilon + 1 + O(\varepsilon) \). The discriminant is \( 36 - 40 + O(\varepsilon) \), hence the inequality holds for small \( \varepsilon \).

4) Balanced \( \|p\|_1 \leq \|u\|_2 \). In a similar way as before, we determine the excesses at the last unbalanced stage and the condition becomes

\[
(1 - 4\varepsilon)(a + b)/\sqrt{2} + 6\varepsilon(a_o + b_o)/\sqrt{2} \leq \sqrt{a^2 + b^2}.
\]

Suppose that \( b_o \leq a_o\sqrt{20\varepsilon} \) (the other case is similar) and let \( x = a/a_o \) and \( y = b/b_o \), thus \( x \geq 1 \). The equation is

\[
(1 - 4\varepsilon)(x + y) + 6\varepsilon + O(\varepsilon^{3/2}) \leq 2x^2 + 2y^2.
\]

First suppose \( y \geq 1 \), thus \( x + y \geq 2 \). Observe that for varying \( x \) and \( y \) the left hand side only depends on \( x + y = z \). Note that \( z \geq 2 \). The smallest value in the right hand side is obtained for \( x = y = z/2 \). In this case, terms without \( \varepsilon \) cancel and the equation becomes

\[
-4z + 6 + O(\sqrt{z}) \leq 0; \text{ which follows from } z \geq 2.
\]

Suppose \( y \leq 1 \). The slope of \( x \) in the left-hand is \( 1 - 4\varepsilon \), and the slope in the right hand side is \( 2x/\sqrt{2x^2 + 2y^2} = \sqrt{2/(1 + (y/x)^2) \geq 1} \) (because \( y \leq 1 \leq x \)). Hence, it suffices to check the equation for \( x = 1 \) (remind that \( x \geq 1 \)): \( y^2/2 + y(2 + 10\varepsilon) + 1 + 8\varepsilon + O(\varepsilon^{3/2}) \geq 0 \). The discriminant is proportional to \( (1 + 5\varepsilon)^2 - (1 + 8\varepsilon)^2 + O(\varepsilon^{3/2}) < 0 \). Hence, \( o \) and \( p \) satisfy all conditions.

A numerical search to find a maximal \( \varepsilon \) satisfying equations \( (7) - (10) \) shows that Theorem 3 holds for \( \beta = 0.491 \).
Asymmetry of online Kolmogorov complexity

Note that in the construction of \( o \) and \( p \), we used a 2-dimensional \( u \) both to determine whether a stage is balanced or unbalanced and to be a linear factor in equations 5 and 6. \( \overline{v}(v, u) \) and \( \overline{P}(v, u) \) are obtained by using a separate 2-dimensional vector \( v = [a, b] \) for the first purpose and \( u \) for the second purpose. Now \( u \) can have any dimension. It remains to show the conditions for induction. Remark that all conditions \( [1]-[10] \) can be written as \( c_1 a + c_2 o + c_3 a_o + c_4 b_o \leq \|[a, b]\|_2 \) where \( c_i \geq 0 \) for \( i = 1, \ldots, 4 \). Consider condition \( 2 \) (\( 1 \) is analogue): for coefficients \( c_1 \ldots c_4 \) corresponding to the balanced or unbalanced case, \( \|\overline{P}(v, u)\|_{1,\infty} = \sum_{i, k} |P_i(u_i, o_i)| \) is bounded by

\[
\|c_1 u_- + c_2 u_+ + c_3 u_{-o} + c_4 u_{+o}\|_{\infty, 1} \leq c_1 \|u_-\|_{\infty, 1} + \cdots + c_4 \|u_{+o}\|_{\infty, 1},
\]

de the inequality follows from homogeneity and the triangle inequality of norms. Let \( v = [a, b] \) and assume that \( \|u_-\|_{\infty, 1} \leq a, \|u_{-o}\|_{\infty, 1} \leq a_o, \|u_+\|_{\infty, 1} \leq b \) and \( \|u_{+o}\|_{\infty, 1} \leq b_o \), then the right hand side is at least \( c_1 a + \cdots + c_4 b_o \leq \|v\|_2 \) by construction of the \( n = 1 \) case. Hence the condition is satisfied.

\section{Appendix: Online prediction of each \( k \)-th bit}

For all \( i, k \), an \( i \) modulo \( k \) semimeasure is the natural generalization of even semimeasures, where condition \( i \) holds if \( |x| = i \mod k \) and \( i \) holds otherwise.

\begin{proposition}
For all \( \epsilon > 0 \), \( k \geq 2 \) and lower semicomputable \( (i \mod k) \)-semimeasures \( Q_i \mod k \) for \( i = 1 \ldots k \), there exist \( \delta > 0 \), a partial computable \( F \), a lower semicomputable semimeasure \( P \), and a sequence \( \omega \) such that for all \( n \)

\[
\left( \prod Q_i \mod k \right)(\omega_{1 \ldots kn}) \leq (1 - \delta)^n P(\omega_{1 \ldots kn})^{k-k\epsilon}
\]

and \( F(\omega_{1 \ldots kn}, \omega_{kn+k}) = \omega_{1 \ldots kn+k} \).
\end{proposition}

\textbf{Proof.} The proof is analogous to the proof above. First consider the game for the case \( n = 1 \). We choose

\[
\delta = 2^{-1 + \frac{\log k}{\epsilon}}
\]

eq 2^{1 - \epsilon} = 2^{1 + \log k} = 2k\delta. \quad \text{Let } 0^k \text{ be the string containing } k \text{ zeros. We choose}

\( (1 - \delta)^1 P(0^k) = \delta \) and wait until \( Q_i \mod k(0^k) > (1 - \delta)^{-1} \) for some \( i = 1, \ldots, k \). Let \( i \) be the first for which this happens and let \( x^i \) be the leftmost string for which the \( i \)-th and the \( k \)-th bit is 1, i.e. \( x^i = 0^{n-1}10^{k-i-1}11 \). Alice’s second (and last) move is \( (1 - \delta)^1 P(x^i) = 1 - 2\delta \).

We have that \( \prod Q_i \mod k \) is \( 1 - 1 + \epsilon = 1 - 2k\delta \), and Bob is the first \( (1 - 2\delta)^k \geq 1 - 2k\delta \), and hence he can win in any particular case, \( Q_i \mod k \) is \( (1 - 2\delta)^k \leq 1 - 2\delta \).

Note that for the string \( x \) on which Alice wins, we have \( x_k \) iff Alice had a second move. Knowing that Alice had a second move, we can compute \( i \). Thus we can compute \( x \) from \( k, x_k \).

For \( n \geq 1 \) we define \( \omega, t_{n,i} \) inductively in a similar way as before. Let the event \( E_{x,i} \) denote whether \( Q_j \mod k(x^i) > (1 - \delta)^{k-\epsilon} \), and this was not detected already for the other cases \( Q_j \mod k \) with \( j \neq i \). Let \( \omega_{n_{k+1 \ldots pn+1}} = x^j \) if an event \( E_{\omega_{1 \ldots kn+1}} \) happens, and otherwise, let \( \omega_{n_{k+1 \ldots pn+1}} = 0^k \). Let \( t_{n+1,i} = t_{n,i} \) if an event \( E_{\omega_{1 \ldots kn+1}} \) happens, and \( t_{n+1,i} = t_{n+1,j} = t_{n,j} \) for all \( j \neq i \), and otherwise let \( t_{n+1,i} = (1 - 2\delta)^{k-2\epsilon} t_{n+1,j} \) if an event \( E_{\omega_{1 \ldots kn+1}} \) happens, and \( t_{n+1,j} = t_{n,j} \) for all \( j \neq i \). For a shorter length \( kn \), we define \( t_{x,i} \) to be \( t_{n,i} \) if at some
point $x$ is an initial segment of a candidate $\omega$ in an approximation of $\omega$ as considered above.

For such $x$ we define $P$ by

$$(1 - \delta)^n P(x) = \left( \prod_{i=1}^{k} t_{x,i} \right)^{1/(k-k\epsilon)}.$$

Now Proposition 9 follows after rescaling $\delta$. $\blacksquare$

### C Appendix: Maximal linear asymmetry

**Proposition 10.** There exist a sequence $\omega$, a lower semicomputable semimeasure $P$ and odd and even online lower semicomputable semimeasures $P_{\text{odd}}$ and $P_{\text{ev}}$ exist such that for all $n$

$$(3/2)^n (M_{\text{odd}}M_{\text{ev}})(\omega_1...\omega_{2n}) \leq P(\omega_1...\omega_{2n}) = (P_{\text{odd}}P_{\text{ev}})(\omega_2\omega_1...\omega_{2n}\omega_{2n-1}).$$

**Proof.** We consider a variant of the game defined before the proof of Proposition 6 where $a, b, c, d$ are values of $2P(x)/3$ (rather than $3P(x)/4$). Alice’s winning strategy is to start with $a = c = 1/9$. As long as she is in winning position, she passes. Suppose at some moment this is no longer true, thus $pr > 1/9$ and $qu > 1/9$ (see Figure 6). Alice’s next (and last) move is $d = 4/9$ if $p \geq q$ and $b = 4/9$ otherwise. Note that $a + b + c + d = 6/9$ and Alice does not violate her restriction. Let $p, q, ..., v$ be Bob’s values at the moment of Alice’s last move, and let $p', q', ..., v'$ denote the limits of Bob’s values. Consider the case $p \geq q$, the other case is analogous. We show that Bob can not win without violating his restriction, i.e. $q'v' > 4/9$ implies $(p' + q')(u' + v') > 1$. Indeed,

$$(p' + q')(u' + v') \geq \left( \sqrt{p'u'} + \sqrt{q'v'} \right)^2. \quad (11)$$

This is Cauchy’s inequality $||\vec{a}|| \cdot ||\vec{b}|| \geq ||\langle \vec{a}, \vec{b} \rangle||$ for $\vec{a} = [\sqrt{p'}, \sqrt{q'}]$ and $\vec{b} = [\sqrt{u'}, \sqrt{v'}]$. Because $u' \geq u$ and $p' \geq p \geq q$, the right-hand is at least

$$\geq \left( \sqrt{p'u'} + \sqrt{q'v'} \right)^2 > \left( \sqrt{\frac{1}{5}} + \sqrt{\frac{4}{5}} \right)^2 = 1. \quad (12)$$

Let $t_0 = 1$. We construct $\omega_{1...2n}$ together with thresholds $t_n$ inductively. For $x$ of length $2n$, let $T_{00}$ and $T_{10}$ at $x$ be the events that $(M_{\text{odd}}M_{\text{ev}})(x00) > t_n/9$ and $(M_{\text{odd}}M_{\text{ev}})(x10) > t_n/9$. Let $p_x$ and $q_x$ denote the values $M_{\text{odd}}(x1)$ and $M_{\text{odd}}(x0)$ at the moment we observe both
Asymmetry of online Kolmogorov complexity

\[ \frac{1}{6} \frac{1}{6} \frac{2}{3} \] \quad \begin{array}{c}
\frac{1}{3} \\ \frac{1}{3} \\ \frac{2}{3}
\end{array} = \begin{array}{c}
\frac{1}{3} \\ \frac{1}{3} \\ \frac{2}{3}
\end{array} \cdot \begin{array}{c}
1 \\ 1 \\ 1
\end{array} \]

\textbf{Figure 7} Decomposition of } P' \text{ defined by Alice's strategy.

\[ T_{00} \text{ and } T_{10} \text{ at } x. \]

\[ \begin{array}{l}
(\omega_{2n+1} \omega_{2n+2}, t_{n+1}) = \\
(00, \ t_n/9) \quad \text{if } T_{00} \text{ does not happen at } \omega_{1...2n} \\
(10, \ t_n/9) \quad \text{if } T_{00} \text{ happens at } \omega_{1...2n} \text{ but not } T_{01}, \\
(11, \ 4t_n/9) \quad \text{if } T_{00}, T_{01} \text{ happen at } \omega_{1...2n} \text{ and } q_x \leq p_x \\
(01, \ 4t_n/9) \quad \text{otherwise.}
\end{array} \]

By induction (and the game above) it follows that \( t_n \geq (M_{\text{odd}} M_{\text{ev}})(\omega_{1...2n}) \). We now define a lower semicomputable semimeasure \( P \) such that

\[ P(\omega_{1...2n}) = (3/4)^n n_{2n} v_n. \]

First, note that \( \omega \) can be approximated as follows: start with \( \omega = 00 \ldots \), if \( T_{00} \) or \( T_{01} \) happen, then bits \( \omega_{2n} \omega_{2n+1} \) are changed accordingly, let all subsequent bits be zero, and run the processes for \( n + 1 \), \( n + 2 \), etc. For each \( n \) and for each \( x \) of length \( 2n \), at most one value \( t_n \) can be associated to \( x \). If this happens, we define \( P(x) = (6/9)^n t_n \) and \( P(x) = 0 \) otherwise. Also note that the two last cases in the definition of \( \omega \) can not happen simultaneously, hence \( \sum \{t_{xb'b'} : b, b' \in \{0,1\}\} \leq 6 t_x / 9 \). This implies \( \sum \{P(xb'b') : b, b' \in \{0,1\}\} \leq P(x) \) and \( P \) is a semimeasure.

It remains to factorize \( \tilde{P}(x_2 x_1 \ldots x_{2n} x_{2n-1}) = P(x_1 x_2 \ldots x_{2n-1} x_{2n}) \) into two online semimeasure \( P_{\text{od}} \) and \( P_{\text{ev}} \). The decomposition for \( P \) is given in figure 7 for \( x \) of length two (the two maximal cases are plotted). This construction can be iterated, (i.e. we obtain \( P_{\text{od}}(xb'b') \) and \( P_{\text{ev}}(xb'b') \) by multiplying the values of figure 7 with \( P(x) \)). In fact, \( P_{\text{od}} \) is computable and \( P_{\text{ev}} \) is non-zero on exactly the same places as \( P \).

Finally, we remark that this result can be generalized for more machines using the generalized Hölder’s inequality, which is in turn a generalisation of the Cauchy-Schwartz inequality: for \( r, s_1, \ldots, s_k \) such that \( \sum_{i=1}^k \frac{1}{s_i} = 1 \), and for vectors \( \vec{u}^1, \ldots, \vec{u}^k \)

\[ \|\vec{u}^1 \ldots \vec{u}^k\|_p \leq \|\vec{u}^1\|_{s_1} \ldots \|\vec{u}^k\|_{s_k}, \]

where \( \vec{a} \vec{b} \) denotes entry wise multiplication.

\textbf{D} \quad \textbf{Appendix: Chain rule for online complexity}

\textbf{Proposition 11.} \( C_{\text{ev}}(x|y) = C_{\text{ev}}(x) + C_{\text{ev}}(y|x) + O (\log |x|) \) and similar for odd complexity.

The proof is similar to the proof of symmetry of information for prefix complexity [19]. A conditional even semimeasure \( P_{\text{ev}}(x|y) \) is defined in the natural way, i.e. a function such that \( P_{\text{ev}}(\cdot|y) \) is an even semimeasure for all \( y \). Note that if even complexity was defined over discrete sets (rather than \( \{0,1\} \)), the conditional variants are simply the cases where
the condition is joined with the first symbol of the string. Hence the general version of
the coding theorem in $[3]$ implies $-\log M_{ev}(x|y) = C_{ev}(x|y) + O(\log C_{ev}(x|y))$. Therefore,
Proposition 11 follows from

\[ M_{ev}(x|x)M_{ev}(y|x, k_x) = \Theta \left( M_{ev}(xy|x) \right) \]

with $k_x = \lceil -\log M_{ev}(x|x) \rceil$.

The proof roughly follows the proof of symmetry of information for prefix complexity,
where it is shown that $m(x|m(x, \lceil -\log m(x) \rceil) = \Theta(m(x,y))$.

**Proof.** Let $m = |x|$. We show that the left hand side exceeds the right within a constant
factor. For this it suffices to construct a lower semicomputable even semimeasure $P_{ev}$ such
that for all $k$ with $2^{-k} \geq M_{ev}(x|m)$ we have

\[ P_{ev}(y|x, k) = \frac{M_{ev}(xy|x)}{2^{-k}}. \]

Indeed, assume $M_{ev}$ is approximated from below such that at each stage $M_{ev}$ is an even
semimeasure. At each stage, take the above function as a definition of $P_{ev}$, as soon as
$M_{ev}(x|x) > 2^{-k}$ do not increase $P_{ev}$ anymore. Clearly $P_{ev}$ is lower semicomputable, the
“freezing” guarantees that $P_{ev}(\varepsilon|x, k) \leq 1$ for all $x$ and $k$ and hence, it is a conditional even
semimeasure.

For the other inequality we construct an even lower semicomputable semimeasure $P_{ev}$
such that if $|z| \geq m$ then

\[ P_{ev}(z|m) \geq \frac{1}{2} M_{ev}(z_{1...m}|m)M_{ev}(z_{m+1...|z|}|z_{1...m}, k), \]

for $k = \lceil -\log M_{ev}(z_{1...m}|m) \rceil$. Our construction of $P_{ev}$ is as follows: if $|z| < m$, then
$P_{ev}(z|m) = M_{ev}(z|m)$ and otherwise

\[ \sum_{k} \left\{ 2^{-k-1} M_{ev}(z_{m+1...|z|}|z_{1...m}, k) : 2^{-k} \leq M_{ev}(z_{1...m}|m) \right\} . \]

Note that for $|z| = m$ we have $P_{ev}(z|m) \leq \sum \{2^{-k-1} : 2^{-k} \leq M_{ev}(z_{1...m}|m) \} \leq M_{ev}(z|m)$
hence $P_{ev}$ is an even semimeasure. Moreover, $P_{ev}$ is lower semicomputable and satisfies the
condition. ◀

In $[1]$, Lemma 12 is combined with Péter Gács’ theorem that $\max\{C(C(x)|x) : |x| = n\} \geq \log n - O(\log \log n)$ to obtain a more involved proof of Theorem 2 in weaker form, i.e.
with a smaller (and machine dependent) linear coefficient.