Федеральное государственное автономное образовательное
учреждение высшего образования
«Национальный исследовательский университет
«Высшая школа экономики»

Факультет компьютерных наук
Основная образовательная программа
Прикладная математика и информатика

КУРСОВАЯ РАБОТА

на тему

Связь монотонной Колмогоровской сложности и априорной вероятности
Monotone Kolmogorov complexity and a priori probability

Выполнил студент группы БПМИ-173, 3 курса,
Урманов Максим Тимурович

Научный руководитель:
Доцент, PhD, Баувенс Бруно Фредерик Л.

Москва, 2020

Contents

1 Abstract

In this work we study the so-called "tree game" which bears a strong connection with the open problem about the gap between the Kolmogorov complexity of a binary string and its a priori probability. We find the winner of the game and their winning strategy based on the parameters of the game and provide the proofs for the established results.

В данной работе рассматривается так называемая «игра на дереве», которая существенным образом связана с открытой проблемой о разрыве между Колмогоровской сложностью бинарной строки и её априорной вероятностью. В работе определяется победитель в игре и находится его выигрышная стратегия в зависимости от параметров игры и доказываются полученные результаты.

2 Motivation

The tree game emerges from an open problem about the relation between learning and compression. In this section we briefly discuss this problem. Later sections do not depend on this section.

Imagine a black box that produces bits one by one. We have no prior information available about the box, except for a string x containing the previously produced bits. Our aim is to predict the next bit that the box might produce.

Since no information is available, nothing can be said definitely, hence, one relies on heuristics to make further predictions. One may consider two basic heuristics:

1. Look for a short program on a Turing machine whose output starts with the already observed string x. Then use the next output bit of this program as prediction.
2. There exists a lower semicomputable semimeasure on infinite binary sequences that is maximal up to a constant factor. Approximate this measure for the strings $x0$ and $x1$ (here ab denotes the concatenation of the strings a and b) and use the bit that makes this measure maximal (0 or 1) as prediction.

How do these heuristics relate to each other? If semimeasures are large for some string x, can we find a short program for x and vice versa? Unlike the first question, the second is easy to formalize. Let $KM(x)$ be the negative logarithm of the maximal semimeasure. The minimal length of the program producing a string x on a machine U is the Kolmogorov complexity $C_U(x)$. There exists a machine that makes this quantity minimal up to additive $O(1)$ terms. We fix such a machine and drop the subscript $_U$. The coding theorem implies that $KM(x)$ is equal to $C(x)$ up to additive terms logarithmic in the length of x. Can this precision be improved?

For this, we need the notion of a monotone Turing machine. This is a machine U such that for any two strings p and q for which both outputs $U(p)$ and $U(pq)$ are defined, either $U(p)$ is a prefix of $U(pq)$ or vice versa, $U(pq)$ is a prefix of $U(p)$. There exists a monotone machine that makes the complexity function $C_U(x)$ minimal among all monotone machines. We fix such a machine U and write $Km(x) = C_U(x)$.

It is not so difficult to show that $Km(x) \geqslant KM(x) - O(1)$. Peter Gacs showed that the equality

$Km(x) = KM(x) + O(1)$ does not hold. The best known lower bound for the difference $Km(x) - KM(x)$ is $\Omega(\log\log n)$ and was proved by Adam Day using a small optimization of Gacs proof. We refer to the book [1] by Shen, Vereshchagin and Uspensky for a nice exposition of this proof.

Still, the proofs are very difficult and raise the following two questions:

1. Does there exist a simple proof that the difference is not bounded by a constant?
2. The best known upper bound on the difference is $\log n + O(\log\log n)$. Is this optimal?

All proofs consider the game that we study here. Our goal is to achieve a better understanding of the game for small trees in the hope to uncover the missing insight that is needed to analyse the game on large trees.

Another speculative application might consider write-once memories (for example, CDROM devices). Imagine we want to store several databases whose entries are produced in chunks. In order to read the database fast, we would like to store each database in a single place. The problem is that we do not know in advance how large each of the databases will become. Suppose we are given a bound on the total size of all databases. How large should the memory be so that we can always store the databases in one piece?

3 Description of the game

Consider an undirected tree T with a fixed root and a real nuber $a \geqslant 1$. Then the game $G_{T,a}$ can be defined in the following way. At every stage of the game every node v of the tree has a non-negative real number $\ell(v)$ assigned to it. Initially, all these numbers are zeros, except for the number assigned to the root, which equals 1 and is fixed throughout the whole game. In addition to this, every node v of the tree has some set $S(v) \subseteq [0, a]$ assigned to it at every moment. When the game starts, all these sets are empty, except for the one corresponding to the root, which is the whole segment $[0, a]$.

Two players called Alice and Bob exchange moves. On her move Alice chooses some nodes of the tree (she may choose from all the nodes except for the root) and increases the values assigned to them in such a way that the value of every node is greater or equal to the sum of values of its children. More formally, we need the condition

$$\forall v \in T \sum_{\substack{u:\ u \text{ is a} \\ \text{child of } v}} \ell(u) \leqslant \ell(v) \tag{1}$$

to hold at any point in the game. This condition implies that the sum of values over the leaves of the tree never exceeds 1.

Bob's objective in the game is to ensure that after every move of his the following condition holds:

$$\text{for every node } v \text{ there exists an interval } I \subseteq S(v) \text{ of length at least } \ell(v). \tag{2}$$

To achieve this, in every turn Bob can do several operations of this kind: take a subset $S \subseteq [0, a]$ and extend the set $S(v)$ for some node v by merging it with the set S. However, Bob must satisfy the following restriction: for every two nodes u, v which share the same parent the sets $S(u)$ and $S(v)$ always have to be disjoint. On each move Bob can do as many of these operations as he wishes, but, naturally, once the set $S(v)$ is merged with some set S, this merge cannot be undone later in the game.

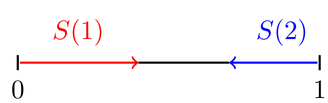
To summarise, Bob's aim in the game is to keep the condition (2) holding true (as it is mentioned above), while Alice's goal is to reach a position in the game where Bob cannot make a move (that is, he cannot sustain the condition (2)). Obviously, the result of the game is determined solely by the parameters of the game, i.e. the tree T and the number a, provided that Alice and Bob play optimally.

In this paper we try to understand which player wins in the case when the tree has depth 1 (which means that all the nodes in the tree are leaves, except for the root) depending on the parameter a, because it is crucial as a first step to studying the game for binary trees, which is by all means more complicated. More precisely, for a fixed tree T of depth 1 we would like to find the smallest value of a such that Bob wins the game $G_{T,a}$. For convenience, we will denote the tree of depth 1 with n leaves by R_n and for a tree T the smallest value of a for which Bob wins in the game $G_{T,a}$ by $a(T)$. Besides, we will use the word "colour" instead of "node" when talking about the problem for trees of depth 1, as in this case the problem has a natural representation in terms of colours (when we merge the set $S(v)$ with some set S, we can think of it as colouring all the points in S in colour v). The colours will be numbered from 1 to n.

4 Known results

To help the reader gain some intuition of the game, we start by discussing some basic known results about the game on trees of depth 1. First of all, let us consider the easiest possible case, which is the game on a tree of depth 1 with 2 leaves (we denote this tree by R_2). Obviously, if $a < 1$, then Bob can do nothing to maintain the condition (2), since Alice can increase the number corresponding to one of the colours by 1 on her very first move and Bob will be unable to respond, as the whole segment $[0, a]$ has length a, which is less than 1. On the other hand, when $a = 1$, Bob has a trivial winning strategy. If 1 and 2 are the numbers of the colours (nodes), he simply keeps $S(1)$ equal to $[0,\ \ell(1)]$ and $S(2)$ to be equal to $[1 - \ell(2),\ 1]$ by extending the segments $S(1)$ and $S(2)$ in the right and left direction respectively when it is needed. It is always possible, because $\ell(1) + (1 - (1 - \ell(2))) = \ell(1) + \ell(2) \leqslant 1$ at every moment in the game.

The picture below shows the strategy of Bob. Arrow directions match the ones in which the corresponding segments are going to be extended by Bob (when necessary).



Therefore, we can make the following proposition.

Proposition 4.1. The smallest value of a such that Bob wins in the game $G_{R_2,a}$ is $a = a(R_2) = 1$.

Another natural question that we can ask right from the start is whether $a(R_n)$ tends to $+\infty$ as n tends to $+\infty$ (for example, if we could find a lower bound on $a(R_n)$ depending on n which would tend to $+\infty$ as n tends to $+\infty$). Surprisingly, the answer here is negative. It turns out that the following theorem holds:

Theorem 4.2. For all n Bob has a winning strategy in the game $G_{R_n,4}$. In other words, for all n we have $a(R_n) \leqslant 4$.

Proof. We present a winning strategy for Bob in the game $G_{R_n,4}$. At every moment for each colour v the set $S(v)$ consists of several disjoint subsegments of $[0,a]$ and the sets $S(v)$ for all colours v together will form a segment $[0,x]$ for some $0 \leqslant x \leqslant 4$, where x is the rightmost endpoint of all the segments allocated so far. Moreover, for each colour v the rightmost of the disjoint segments of this colour will be called *active*, while all the other segments of this colour will be called *inactive*. Bob's strategy will be as follows: if after Alice's last move some colour v has a value $\ell(v)$ which is greater than the length of its active segment, Bob allocates a new active segment $[x,\ x+2\ell(v)]$ for this colour (that is, he merges $S(v)$ with this new segment) and updates the value of x by setting $x := x + 2\ell(v)$. On his move Bob repeats this procedure for as many colours as necessary (it is clear that each of the colours needs at most one such procedure). This sums up the whole strategy. Now it remains to prove that the value of x will never become greater than 4. It is quite easy to see this, because the total amount of space occupied by Bob specifically for the colour v at any moment is upper-bounded by the sum

$$\sum_{i=-1}^{k-2} \ell(v) \cdot 2^{-i}, \tag{3}$$

where k is the number of the above mentioned procedures performed on colour v, because the length of the current active segment of v is at most $2\ell(v)$, while the previous active segment of v is at most half as big as the current active segment and the segment which was active before the previous one is at most half as big as the previous one and thus is at most $1/4$ as big as the currently active segment and so on, which gives the upper bound (3). Now we can note that the sum in (3) is itself upper-bounded by the sum

$$\sum_{i=-1}^{+\infty} \ell(v) \cdot 2^{-i} = \ell(v) \cdot (2 + 1 + 1/2 + 1/4 + \ldots) = 4\ell(v). \tag{4}$$

Since the sum of the values $\ell(v)$ over the colours never exceeds 1, we get the desired result. $\square$

This result immediately gives the following corollary.

Corollary 4.3. In the game on the binary tree of depth n (i.e. the binary tree with 2^n leaves) Bob wins for $a = 4n$.

The proof is obvious: Bob simply plays an independent game on each level of the tree according to the strategy described in the proof of lemma 4.2, thus consuming the space of at most $4 \cdot n$.

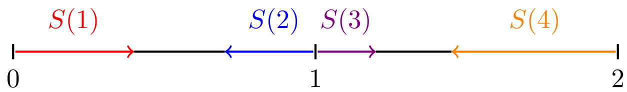
5 Analysis of the games on R_3 and R_4

The first question which arises after understanding everything about the game on R_2 is naturally the one about what happens in the game on R_3. We can easily see that $a = 2$ is enough for Bob to seize the victory in the games on R_3 and R_4. Let us prove it as a lemma.

Lemma 5.1. Bob has a winning strategy in the game $G_{R_4,2}$.

Proof. The strategy is quite straightforward: split the segment $[0, 2]$ into segments $[0, 1]$ and $[1, 2]$, then independently play a game on R_2 for colours 1 and 2 on the segment $[0, 1]$ and a game on R_2 for colours 3 and 4 on the segment $[1, 2]$. $\square$

This strategy is depicted in the figure below.



The given strategy may seem to be too simple to be optimal for R_3 and R_4. However, it turns out that $a(R_4) = a(R_3) = 2$. We are going to prove this result shortly, but first we need to introduce a new technique.

Note that in any tree game there are essentially two types of moves that Bob can perform, namely, *allocations* and *extensions*. Moves of the first type involve allocating an entirely new segment for some colour v, this segment being disjoint with all the segments previously allocated for this colour. Moves of the second type imply extending an already existing segment of some colour v. Clearly, the moves of the first type are, in a way, more "expensive" than those of the second type, because when $\ell(v)$ is increased by Alice it's cheaper (merely in terms of space consumption, we are not talking about optimal strategies yet) for Bob to extend an existing segment of colour v so that its length equals $\ell(v)$ (it is not optimal to extend the segment to a greater length than the precise value of $\ell(v)$, as Bob can do it mentally if he wishes), rather than to allocate an entirely new segment of length $\ell(v)$. This basic intuition gives us a hint at what Alice has to do to make Bob consume as much space as possible. Namely, she needs to force Bob to make a lot of allocation moves. Our strategy will be based on this idea. To formalise the processes of extension and allocation we need to bring in several definitions.

Definition 5.2. Consider a finite set S of coloured pairwise disjoint subsegments of $[0, a]$. For a segment $[p, q] \in S$ of colour v consider the points

$$p' := \inf\{x : \ \forall t \in [x, p] \ t \text{ has colour } v \text{ or does not have any colour at all}\}$$

and

$$q' := \sup\{x : \ \forall t \in [q, x] \ t \text{ has colour } v \text{ or does not have any colour at all}\}.$$

To put it simply, $[p', q']$ is the largest segment to which $[p, q]$ can be extended. Then we shall call the segment $[p', q']$ the *extension* of the segment $[p, q]$ and the number $\mathsf{EC}([p, q]) := (q' - q) + (p - p')$ the *extension capacity* of the segment $[p, q]$.

Definition 5.3. Consider a finite set S of coloured pairwise disjoint subsegments of $[0, a]$. For a colour v we define the *extension capacity* of the colour v (denoted by $\mathsf{EC}(v)$) as the maximal extension capacity among the segments of colour v. Mathematically, the definition is given by the following formula:

$$\mathsf{EC}(v) := \max_{\substack{s \in S: \\ s \text{ has colour } v}} \mathsf{EC}(s).$$

We will proceed by proving a trivial, yet really useful lemma for our goal.

Lemma 5.4. Consider an arbitrary position in the game $G_{R_3, 2-c}$, where c is an arbitrary positive real number. Suppose that $\ell(v) > 0$ for all $v \in \{1, 2, 3\}$. Denote by XB the total space consumed by Bob. Then there is a colour $v \in \{1, 2, 3\}$ such that

$$\mathsf{EC}(v) \leqslant 1 - \frac{c}{2} - \frac{XB}{2}.$$

Proof. For each colour let us take the segment of this colour with the largest extension capacity. These three segments are pairwise disjoint, hence one of them lies between the other two. Denote the two outer segments by s_{left} and s_{right}. Note that the extensions of these segments are disjoint as well (since a segment of the third colour lies between them). This means that

$$\mathsf{EC}(s_{\text{left}}) + \mathsf{EC}(s_{\text{right}}) \leqslant 2 - c - XB,$$

therefore,

$$\min(\mathsf{EC}(s_{\text{left}}), \mathsf{EC}(s_{\text{right}})) \leqslant \frac{2 - c - XB}{2} = 1 - \frac{c}{2} - \frac{XB}{2}.$$

It remains to set v to be the colour of the segment with the lowest extension capacity between s_{left} and s_{right}. ◻

Now we are ready to prove the main lemma.

Lemma 5.5. Alice has a winning strategy in the game $G_{R_3, 2-c}$, where c is an arbitrary positive number.

Proof. Our strategy for Alice will be as follows. On the first move we set the values $\ell(1), \ell(2)$ and $\ell(3)$ to be equal to $1/n$ for some large positive integer n. We will determine the value of n later. Starting from the second move, we will split the game into *phases*. In each phase Alice will make a series of moves which will end with Bob responding with an *allocation* move. As soon as Bob performs an *allocation* move, the phase ends and we start the next phase. Now we are going to describe Alice's strategy in any phase in details. At the start of the phase Alice picks the colour v with the smallest extension capacity. Then she makes a series of the following moves: increase the value $\ell(v)$ by $1/n^2$. By saying "series" here we imply that Alice makes a move (one increase of $\ell(v)$ by $1/n^2$), then waits for Bob's response. If Bob performs an *allocation* move on colour v, the phase ends. Otherwise Bob makes an *extension* (because we have already observed that the largest segment of colour v can always have length precisely $\ell(v)$ in Bob's optimal strategy) and Alice makes the next increase, and so on. Note that the colour v is selected at the start of the phase and does not change throughout the whole phase. Certainly, it is not obvious that any phase will end at some point. However, if we prove the following two claims:

1. any phase in our strategy will reach an ending
2. after any successful phase Alice will be able to complete a new one,

we will prove the lemma. This is due to the fact that the game can last only for a finite number of moves, namely, at most $2/(1/n) = 2n$ phases, since each colour v has a value $\ell(v) \geqslant 1/n$ at every moment of the game and thus every phase ends with Bob adding at least $1/n$ to his total consumption. So, to finish the proof of the lemma, it suffices to prove the two claims written above. We are going to do this by induction on k — the number of the current phase.

We start with the base case $k = 1$. At the beginning of the first phase Bob's total consumption is simply $3/n$. Then, by lemma 5.4, there is a colour v such that

$$\mathsf{EC}(v) \leqslant \frac{2 - c - 3/n}{2} = 1 - \frac{c}{2} - \frac{3}{2n}.$$

Thus, to complete this phase, Alice will need to increase the value of l of the the colour with the smallest extension capacity by at most $1 - \frac{c}{2} - \frac{3}{2n} + \frac{1}{n^2}$ (the additional $1/n^2$ is for the last step of the phase).
At the same time, Alice's free space at this moment equals $1 - 3/n$. If n is small enough, we have $c > 3/n + 2/n^2$, hence

$$1 - \frac{c}{2} - \frac{3}{2n} + \frac{1}{n^2} < 1 - \frac{3}{2n} - \frac{1}{n^2} - \frac{3}{2n} + \frac{1}{n^2} = 1 - \frac{3}{n},$$

which means that Alice can complete the first phase. If we denote the space consumed by Alice on the first phase by d_1, then Bob's consumption on the first phase is no less than $2d_1 - 1/n^2 + \ell(v)$, where $\ell(v)$ is the value before the start of the phase. This is because Bob spends twice as much space as Alice does during the phase (the phase ends with Bob making an allocation move) plus at least the additional $\ell(v)$ and minus Alice's last move, which is $1/n^2$, as is spent by both Alice and Bob once. Since $\ell(v) \geqslant 1/n$ for all v at every moment, we have a following lower bound on the total space XB_1 consumed by Bob after the first phase:

$$XB_1 \geqslant \frac{3}{n} + 2d_1 - \frac{1}{n^2} + \frac{1}{n}.$$

We can actually generalise this bound to k phases. The total consumption XB_k of Bob after k phases can be lower bounded as follows:

$$XB_k \geqslant \frac{3}{n} + 2\sum_{i=1}^{k} d_i - \frac{k}{n^2} + \frac{k}{n}, \tag{5}$$

where d_i is the consumption of Alice on the i-th phase. Recalling that the number of phases can be at most $2n$, we can eliminate the term k/n^2 by upper-bounding it by $2n/n^2 = 2/n$. Hence our final lower bound:

$$XB_k \geqslant \frac{3}{n} + 2\sum_{i=1}^{k} d_i - \frac{2}{n} + \frac{k}{n} = 2\sum_{i=1}^{k} d_i + \frac{k+1}{n}, \tag{6}$$

whereas the total consumption XA_k of Alice after k phases simply amounts to $\frac{3}{n} + \sum_{i=1}^{k} d_i$.

Now it remains to show the induction step. Suppose that Alice has successfully completed k phases. Let us show that she can complete one more phase. After the first k phases Bob's total consumption XB_k is lower-bounded by the right-hand side of (6). By lemma 5.4, the colour with the smallest extension capacity has an extension capacity of at most

$$1 - \frac{c}{2} - \frac{XB_k}{2} \leqslant 1 - \frac{c}{2} - \frac{1}{2}\left(2\sum_{i=1}^{k} d_i + \frac{k+1}{n}\right) = 1 - \frac{c}{2} - \sum_{i=1}^{k} d_i - \frac{k+1}{2n},$$

while Alice's current free space is $1 - \frac{3}{n} - \sum_{i=1}^{k} d_i$. So, to prove that Alice can complete a new phase, it suffices to show that

$$\left(1 - \frac{c}{2} - \sum_{i=1}^{k} d_i - \frac{k+1}{2n}\right) + \frac{1}{n^2} < 1 - \frac{3}{n} - \sum_{i=1}^{k} d_i.$$

This inequality is equivalent to

$$\frac{c}{2} + \frac{k+1}{2n} > \frac{3}{n} + \frac{1}{n^2}.$$

For this to hold true, we just need to require $c > 6/n + 2/n^2$, which is a little stronger than our requirement for the first phase (which was $c > 3/n + 2/n^2$), but is still fairly doable by taking n to be big enough. Thus we have shown that Alice can complete as many phases as she wishes, which means that she wins the game. $\square$

Lemmas 5.1 and 5.5 together give a proof of the following theorem.

Theorem 5.6. $a(R_3) = a(R_4) = 2$.

Proof. For all n we obviously have $a(R_n) \leqslant a(R_{n+1})$. Lemma 5.1 gives the upper bound $a(R_4) \leqslant 2$, while lemma 5.5 gives the lower bound $a(R_3) \geqslant 2$. Therefore, we get

$$2 \leqslant a(R_3) \leqslant a(R_4) \leqslant 2,$$

which yields $a(R_3) = a(R_4) = 2$. $\square$

We have now delivered a complete analysis of the games $G_{R_3,a}$ and $G_{R_4,a}$, which was the aim of this section. However, we can also see here how games on trees of depth 1 help to understand games on binary trees of bigger depth by deriving the following corollary from theorem 5.6.

Corollary 5.7. Let B_2 be a binary tree of depth 2 (i.e. the one with $2^2 = 4$ leaves). Then $a(B_2) = 2$.

Proof. The values $\ell(v)$ of all the leaves v always sum to at most 1 and their corresponding sets $S(v)$ are required to be pairwise disjoint (otherwise there would be a problem on the upper level), so, the game on the leaves only is equivalent to the game on R_4. Therefore, we have a lower bound $a(B_2) \geqslant a(R_4) = 2$. On the other hand, there is a trivial strategy which allows Bob to win in the game $G_{B_2,2}$. Let us number the nodes of B_2 as follows: the nodes on the first level have numbers 1 and 2 from left to right and the nodes on the second level have numbers $3, 4, 5, 6$ also from left to right. Thus nodes $3, 4$ are the children of node 1 and nodes $5, 6$ are the children of node 2. Then Bob sets $S(1) = [0, 1]$ and $S(2) = [1, 2]$ for the whole game. As for the nodes $3, 4, 5, 6$, Bob maintains their corresponding sets in the following manner: $S(3) = [0,\ \ell(3)]$, $S(4) = [1 - \ell(4),\ 1]$, $S(5) = [1,\ 1 + \ell(5)]$, $S(6) = [2 - \ell(6),\ 2]$ by means of extension. Since $\ell(3)+\ell(4)+\ell(5)+\ell(6) \leqslant 1$ at every moment, Bob is always capable of performing the necessary extensions. Therefore, we have $2 \leqslant a(B_2) \leqslant 2$, which concludes the proof. $\square$

6 Lower and upper bounds for the games on R_5 and R_6

Unfortunately, we have not been able to fully comprehend the games $G_{R_5,a}$ and $G_{R_6,a}$, as they have proved to be more complex than the games on R_3 and R_4. However, we still establish some bounds on the values of $a(R_5)$ and $a(R_6)$. To prove them, we first need to see how lemma 5.4 (about the existence of a colour with low extension capacity) changes in case of R_n instead of R_3. We can generalise lemma 5.4 in the following way.

Lemma 6.1. Consider an arbitrary position in the game $G_{R_n,a}$. Suppose that $\ell(v) > 0$ for all $v \in \{1, \ldots, n\}$. Denote by XB the total space consumed by Bob. Then there is a colour $v \in \{1, \ldots, n\}$

such that

$$\mathsf{EC}(v) \leqslant \frac{a - XB}{\lceil n/2 \rceil}.$$

Proof. For each colour let us take the segment of this colour with the largest extension capacity. These n segments are pairwise disjoint, so we can number them from 1 to n as they lie on $[0, a]$ from left to right. Note that the extensions of the segments with odd numbers are pairwise disjoint as well. Since the total number of odd segments is $\lceil n/2 \rceil$ and the sum of their lengths is at most $a - XB$, we have

$$\min\{\mathsf{EC}(v) \mid v \in \{1, \dots, n\}\} \leqslant \frac{a - XB}{\lceil n/2 \rceil}.$$

□

With this estimate in mind, we can prove a non-trivial result for the game on R_5.

Theorem 6.2. $a(R_5) \geqslant 2 + \dfrac{2}{7 \cdot 3^{10}}$.

Proof. Let us set $a := 2 + c$ for some positive constant c which is less than 1. We will analyse the game $G_{R_5,2+c}$ and in the end choose the value of c that will guarantee the win for Alice according to a sufficient condition that we derive as a result of our analysis. Our strategy for Alice will be similar to the one for the game on R_3. Namely, on the first move Alice sets the values $\ell(v)$ for all colours v to be equal to $1/n$, where n is a large positive number (not necessarily an integer), the exact value of which we will determine later. Then Alice performs the phases in exactly the same way as she did in the game on R_4, i.e. by taking the colour v with the smallest extension capacity and increasing the value of $\ell(v)$ via a series of additions of $1/n^2$. Then the formula for Alice's consumption after k phases (denoted by XA_k) almost remains the same as it was in our proof of lemma 5.5:

$$XA_k = \frac{5}{n} + \sum_{i=1}^{k} d_i,$$

where d_i is the amount of space spent by her on the i-th phase. Analogously, Bob's consumption can be lower bounded as

$$XB_k \geqslant \frac{5}{n} + 2\sum_{i=1}^{k} d_i - \frac{k}{n^2} + \frac{k}{n}.$$

Since the number of phases in the game cannot exceed $(2 + c)/(1/n) = (2 + c)n < 3n$, we have $k/n^2 < 3n/n^2 = 3/n$, obtaining the following lower bound:

$$XB_k \geqslant \frac{5 + k - 3}{n} + 2\sum_{i=1}^{k} d_i = \frac{2 + k}{n} + 2\sum_{i=1}^{k} d_i. \tag{7}$$

By lemma 6.1, the colour v with the smallest extension capacity has an extension capacity

$$\mathsf{EC}(v) \leqslant \frac{2 + c - \left(\frac{2+k}{n} + 2\sum\limits_{i=1}^{k} d_i\right)}{\lceil 5/2 \rceil} = \frac{2 + c}{3} - \frac{2 + k}{3n} - \frac{2}{3}\sum_{i=1}^{k} d_i, \tag{8}$$

while Alice's remaining space simply equals

$$1 - XA_k = 1 - \frac{5}{n} - \sum_{i=1}^{k} d_i. \tag{9}$$

These estimates allow us to write a sufficient condition for Alice to be able to complete phase $k+1$. This condition requires the right-hand side of (9) to be greater than the right-hand side of (8). In the proof of lemma 5.5 we wrote a similar sufficient condition, but required Alice's free space to be greater than the upper bound on Bob's free space by at least $1/n^2$. However, this additional gap of size $1/n^2$ is not really needed, as Alice can make the last increase in the phase (if the phase takes the maximal possible number of moves) to be infinitesimally small. So, our sufficient condition will be as follows:

$$1 - \frac{5}{n} - \sum_{i=1}^{k} d_i > \frac{2+c}{3} - \frac{2+k}{3n} - \frac{2}{3}\sum_{i=1}^{k} d_i,$$

which is equivalent to

$$1 - c > \frac{13-k}{n} + \sum_{i=1}^{k} d_i. \tag{10}$$

Now we will take c to be equal to $1/n$ and write the condition (10) in terms of $1/n$ only. We get

$$1 > \frac{14-k}{n} + \sum_{i=1}^{k} d_i. \tag{11}$$

We can note that if k is greater or equal to 10, then the condition holds automatically without any requirements on n, since the sum $\sum\limits_{i=1}^{k} d_i$ cannot be greater than $1 - 5/n$ by the definition of d_i (remember that Alice spends $5/n$ before the first phase). It means that for Alice it is enough to complete the first 10 phases to win. Yet, to show how Alice is able to do this, we need to prove an additional lemma. Let us fix some phase k in the strategy described above for the game $G_{R_5,2+c}$. Recall that the total space spent by Alice is $XA_k = 5/n + \sum\limits_{i=1}^{k} d_i$. For convenience, let us denote the sum $\sum\limits_{i=1}^{k} d_i$ by D_k.

Lemma 6.3. Suppose that Alice successfully completes the phase number $k+1$. Then the following inequality holds:

$$1 - D_{k+1} \geqslant \frac{1}{3}(1 - D_k) - \frac{c}{3}. \tag{12}$$

The intuition behind this inequality is quite plain. It basically states that the difference between 1 and D_k decreases at most exponentially with an additional term $-c/3$, which, however, will prove not to be crucial.

Proof. From (8) we know that on the phase number $k+1$ Alice spends at most

$$\frac{2+c}{3} - \frac{2+k}{3n} - \frac{2}{3}D_k \leqslant \frac{2+c}{3} - \frac{2}{3}D_k$$

space. Hence, we have

$$D_{k+1} \leqslant D_k + \frac{2+c}{3} - \frac{2}{3}D_k = \frac{2+c}{3} + \frac{1}{3}D_k,$$

which is equivalent to

$$1 - D_{k+1} \geqslant 1 - \left(\frac{2+c}{3} + \frac{1}{3}D_k\right) = \frac{1-c}{3} - \frac{1}{3}D_k = \frac{1}{3}(1-D_k) - \frac{c}{3}.$$

□

Let us get back to the proof of theorem 6.2. Using lemma 6.3, we can prove that

$$1 - D_k \geqslant 3^{-k} - c\sum_{i=1}^{k} 3^{-i} > 3^{-k} + c\sum_{i=1}^{\infty} 3^{-i} = 3^{-k} + \frac{c}{2} \tag{13}$$

by induction on k. Moving $D_k = \sum_{i=1}^{k} d_i$ to the left-hand side of the sufficient condition (11), then replacing the left-hand side by the right-hand side of (13) (which is a lower bound) and, finally, remembering that we have set c to be equal to $1/n$, we get a new sufficient condition (we can also change the comparison sign from $>$ to $\geqslant$, because the lower bound is strict):

$$3^{-k} + \frac{1}{2n} \geqslant \frac{14-k}{n}.$$

Transfering all the terms with k to the right-hand side and all terms with n to the left-hand side, we can rewrite this inequality in the following form:

$$n \geqslant \frac{1}{2}(27-2k)\cdot 3^k, \tag{14}$$

which has to hold for all $k \in \{0, \ldots, 10\}$. Applying some calculus, we get that the right-hand side of (14) assumes its maximum value on $[0, 10]$ at the point $k = 10$ and the maximum itself equals $7 \cdot 3^{10}/2$. Hence, setting $n = 7 \cdot 3^{10}/2$ guarantees the win for Alice. Recalling that $c = 1/n$, we get $c = \dfrac{2}{7 \cdot 3^{10}}$. The proof of theorem 6.2 is complete. □

Remark. Note that the constant c in the proof of theorem 6.2 is not optimal. Using the same technique, a proof for a better constant can be constructed. We do not venture to maximise the constant in the proof above for the sake of simplicity and due to the fact that we believe that it is impossible to prove the exact lower bound on the real value of $a(R_5)$ with this approach. So, the main purpose of introducing theorem 6.2 is in fact showing that $a(R_5) > a(R_4)$.

We now prove an upper bound on $a(R_6)$ by constructing a certain strategy for Bob.

Theorem 6.4. $a(R_6) \leqslant 2 + \dfrac{1}{3}$.

Proof. We will explain the strategy that Bob can apply to win the game $G_{R_6, 2+\frac{1}{3}}$. We start by splitting the segment $\left[0,\ 2+\frac{1}{3}\right] = \left[0,\ \frac{7}{3}\right]$ into 3 subsegments as follows: $\left[0,\ \frac{2}{3}\right],\ \left[\frac{2}{3},\ \frac{4}{3}\right],\ \left[\frac{4}{3},\ \frac{7}{3}\right]$. The lengths of the subsegments are $2/3,\ 2/3$ and 1 respectively. After that Bob is going to maintain the sets $S(v)$ for all colours v in the following way: $S(1) = [0,\ \ell(1)],\ S(2) = [2/3-\ell(2),\ 2/3],\ S(3) = [2/3,\ 2/3+\ell(3)],\ S(4) = [4/3-\ell(4),\ 4/3],\ S(5) = [4/3,\ 4/3+\ell(5)],\ S(6) = [7/3-\ell(6),\ 7/3]$.

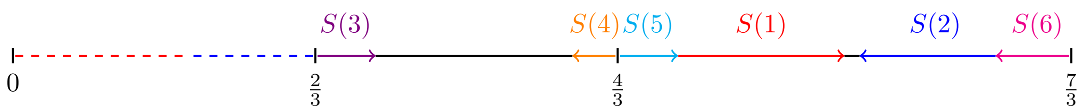
The figure below depicts the strategy so far.

However, this strategy does not work yet, because the sum $\ell(1)+\ell(2)$ may actually become greater than $2/3$. The same problem can happen with colours 3 and 4 ($\ell(3)+\ell(4)$ becomes greater than $2/3$). But note that the problem can happen with at most one of these two pairs of colours. Without loss of generality, suppose that the problem occurs with colours 1 and 2. In other words, the sum $\ell(1)+\ell(2)$ becomes greater than $2/3$ at some moment in the game. In this case, Bob allocates new segments for these two colours. These segments are going to be

$$[4/3+\ell(5),\ 4/3+\ell(5)+\ell(1)] \text{ for colour 1,} \qquad [7/3-\ell(6)-\ell(2),\ 7/3-\ell(6)] \text{ for colour 2.}$$

Note that this move is correct, since $\ell(1)+\ell(2)+\ell(5)+\ell(6) \leqslant 1$ at every moment of the game.

The new state of the game is shown in the picture below. The dashed parts are the ones not being used by Bob anymore.



As we can see from the illustration, now Bob also needs to allocate new segments for colours 5 and 6. For this purpose Bob is going to use the segment $H = [2/3+\ell(3),\ 4/3-\ell(4)]$, which is free now. To show how to do this, first we need to do a small calculation. Denote the sum $\ell(3)+\ell(4)+\ell(5)+\ell(6)$ by σ. Note that $\sigma \leqslant 1/3$, since $\ell(1)+\ell(2) \geqslant 2/3$. We would like Bob to place the new segments for colours 5 and 6 in the middle of the segment H next to each other so that the free space that is left on this segment consists of two parts of equal length: the first part between the segments of colours 3 and 5 and the second part between the segments of colours 6 and 4. This means that each of the two parts will have length $(2/3-\sigma)/2 = 1/3-\sigma/2$ and the exact coordinates of the newly allocated segments will be

$$[2/3+\ell(3)+(1/3-\sigma/2),\ 2/3+\ell(3)+(1/3-\sigma/2)+\ell(5)] \text{ for colour 5,}$$

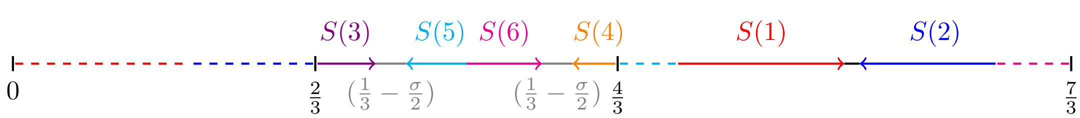
$$[4/3-\ell(4)-(1/3-\sigma/2)-\ell(6),\ 4/3-\ell(4)-(1/3-\sigma/2)] \text{ for colour 6.}$$

Note that the right endpoint of the new segment of colour 5 coincides with the left endpoint of the new segment of colour 6, since

$$2/3+\ell(3)+(1/3-\sigma/2)+\ell(5)+\ell(4)+(1/3-\sigma/2)+\ell(6) =$$

$$= (2/3+1/3+1/3)+\ell(3)+\ell(4)+\ell(5)+\ell(6)-\sigma = 4/3.$$

The new configuration of the game is displayed below. Gray segments are the parts of the free space on H described above, both of length $1/3-\sigma/2$ each.



We claim that from this moment Bob can successfully defend using only extension moves. The only point which we need to verify for this is that $1/3 - \sigma/2$ is greater or equal to the remaining space of Alice. The latter equals precisely $1 - \sum_{i=1}^{6} \ell(i) = (1 - \ell(1) - \ell(2)) - \sigma \leqslant (1 - 2/3) - \sigma = 1/3 - \sigma \leqslant 1/3 - \sigma/2$, which confirms our claim. Hence, our strategy works. □

Theorems 6.2 and 6.4 shed some light on the games on R_5 and R_6. We have established that

$$2 + \frac{2}{7 \cdot 3^{10}} \leqslant a(R_5) \leqslant a(R_6) \leqslant 2 + \frac{1}{3}.$$

Yet, we still lack understanding even of these games, which are played on rather small trees. Our conjecture is that in fact $a(R_5) = a(R_6) = 2 + 1/3$, but more research is required to validate it.

7 A family of strategies for Bob on R_n for $n \leqslant 12$

In this section we describe a family of non-trivial strategies for Bob in the games with up to $n = 12$ colours. We call this set of strategies a *family* because all of them are based on the same idea and differ only slightly implementation-wise depending on the number of colours at play. The idea is, again, quite straightforward. Recall that in the game on R_6 Bob has one free segment of length 1 (which it is absolutely necessary for Bob to have, as otherwise Alice can win by setting $\ell(v) \approx 1$ for an arbitrary colour v) and two segments of length $2/3$ each. There are two such segments essentially because $2 = \lceil 5/2 \rceil - 1 = \lceil 6/2 \rceil - 1$, where $\lceil 5/2 \rceil$ is the number of pairs of colours and we subtract 1 because it corresponds to the single segment of length 1. Bob's game plan involves several allocation moves and he makes a heavy use of the fact that at most one "problem" can ever emerge. The success of this strategy prompts us to try to extend this idea to larger values of n. For example, in case of $n = 7$ or 8 we can reserve one segment of length 1 and take the other $\lceil 7/2 \rceil - 1 = \lceil 8/2 \rceil - 1 = 3$ segments to be of some equal length $x > 1/2$. The fact that there are 3 segments of length x allows Bob to perform 3 reallocations insted of 2 (as we did for R_6), which allows x to be less than $2/3$. We are going to state the following theorem.

Theorem 7.1. There is winning strategy for Bob in the game on R_n which requires $a = (\lceil n/2 \rceil - 1)x + 1$ space, where x is the solution of the equation

$$x + \frac{x}{2} + \frac{x}{4} + \ldots + \frac{x}{2^{\lceil n/2 \rceil - 2}} = 1. \tag{15}$$

Proof. We will describe the strategy for Bob and prove its correctness, obtaining the equality (15) as a result. The strategy for R_n is similar to the one for R_6 from the previous section. We start by splitting the segment $[0, a] = [0,\ (\lceil n/2 \rceil - 1) \cdot x + 1]$ into the segments $r_i := [i \cdot x,\ (i+1) \cdot x]$ for $0 \leqslant i \leqslant \lceil n/2 \rceil - 1$ of length x and the last segment $[\lceil n/2 \rceil x,\ \lceil n/2 \rceil x + 1]$ of length 1. We initialise the sets $S(v)$ for colours $v = 1, \ldots, n$ in the same way as for R_6: for $1 \leqslant i \leqslant \lceil n/2 \rceil - 1$ the sets $S(2i-1)$ and $S(2i)$ are maintained in the form $[(i-1) \cdot x,\ (i-1) \cdot x + \ell(2i-1)]$ and $[i \cdot x - \ell(2i),\ i \cdot x]$ respectively. The last one or two colours are placed on the remaining segment of length 1 in the same way (at the ends). At some point a problem can emerge: for some $1 \leqslant i \leqslant \lceil n/2 \rceil - 1$ the sum $\ell(2i-1) + \ell(2i)$ becomes greater than x. When this happens, Bob reallocates these two colours to the ends of the remaining space on the last segment and reallocates the colours $n - 1$ and n (we can consider the harder case, which is when n is even) to the center of one of the rest of the segments of length x so that the remaining space on this segment is divided into two equal parts (exactly as in the strategy for R_6). Note that Alice cannot "attack" both of these parts (and by "attack" here we mean forcing Bob to make an allocation move), so the worst case scenario for Bob is when she manages to attack one of these two parts. Should this happen, Bob is to reallocate the two colours which share the attacked space to the center of some other segment of length x (again, dividing its free space into two equal parts) which has not been under attack yet. The fact that $x > 1/2$, which is equivalent to $2x > 1$, guarantees that such reallocation is always a correct move. According to our strategy, Bob simply needs to keep reallocating when it is necessary.

To prove that the strategy is correct, we would like to show that the number of such reallocations will never exceed the number of segments of length x minus one, namely, $\lceil n/2 \rceil - 2$. To do this, we begin with the following lemma.

Lemma 7.2. Let $\ell_0(v)$ be the value of $\ell(v)$ right before the attack on the segment where the colour v was initially located (if an attack never happens, ℓ_0 is undefined); let v_{2k-1} and v_{2k} be the colours initially placed on the k-th attacked segment and define $L(k) := L_0 + \sum_{i=1}^{k} \Big(\ell_0(v_{2i-1}) + \ell_0(v_{2i})\Big)$, where L_0 is the total space consumed by Alice on the two colours which were initially on the segment of length 1. Finally, let $\Sigma(k)$ be the total space spent by Alice to perform the first k attacks. Then the following inequality holds:

$$\Sigma(k+1) - \Sigma(k) \geqslant \frac{x - L(k) - \Sigma(k)}{2}. \tag{16}$$

Proof. The total space of the four colours on the attacked segment number $k+1$ is at most $\Sigma(k) + L(k)$. Thus, its free space is at least $x - \Sigma(k) - L(k)$. Alice will have to spend half of this space to perform a new attack, hence the result. $\square$

Lemma 7.2 implies an important inequality, which we also state as a lemma.

Lemma 7.3. The following inequality holds for all k:

$$\Sigma(k) \geqslant \sum_{i=1}^{k} \frac{x}{2^i} - \sum_{i=1}^{k} \frac{L(i)}{2^{k-i+1}}. \tag{17}$$

Proof. We will proceed by induction on k. For $k = 1$ we have $\Sigma(1) \geqslant \dfrac{x - L(1)}{2}$ by lemma 7.2, because $\Sigma(0) = 0$. Let us prove the induction step. By lemma 7.2,

$$\Sigma(k+1) \geqslant \Sigma(k) + \frac{x - L(k) - \Sigma(k)}{2} = \frac{x}{2} + \frac{\Sigma(k)}{2} - \frac{L(k)}{2}. \tag{18}$$

At the same time, we know that (17) holds for k, hence

$$\Sigma(k) \geqslant \sum_{i=1}^{k} \frac{x}{2^i} - \sum_{i=1}^{k} \frac{L(i)}{2^{k-i+1}}.$$

Plugging the right-hand side of this inequality instead of x into the right-hand side of (18), we get

$$\Sigma(k+1) \geqslant \frac{x}{2} + \sum_{i=1}^{k} \frac{x}{2^{i+1}} - \sum_{i=1}^{k} \frac{L(i)}{2^{(k+1)-i+1}} - \frac{L(k)}{2} = \sum_{i=1}^{k+1} \frac{x}{2^i} - \sum_{i=1}^{k+1} \frac{L(i)}{2^{k-i+1}},$$

which completes the induction step. $\square$

We derive the proof of the theorem from lemma 7.3 as follows. The total space consumed by Alice after she performs $\lceil n/2\rceil - 2$ attacks is at least $x + L(\lceil n/2\rceil - 2) + \Sigma(\lceil n/2\rceil - 2)$. Applying lemma 7.3 to the term $\Sigma(\lceil n/2\rceil - 2)$, we get that Alice's total consumption is at least

$$x + L(\lceil n/2\rceil - 2) + \sum_{i=1}^{\lceil n/2\rceil-2} \frac{x}{2^i} - \sum_{i=1}^{\lceil n/2\rceil-2} \frac{L(i)}{2^{\lceil n/2\rceil-2-i+1}} =$$

$$= \left(x + \frac{x}{2} + \frac{x}{4} + \ldots + \frac{x}{2^{\lceil n/2\rceil-2}}\right) + \left(L(\lceil n/2\rceil - 2) - \sum_{i=1}^{\lceil n/2\rceil-2} \frac{L(i)}{2^{\lceil n/2\rceil-1-i}}\right).$$

Note that for all i we have $L(i) \geqslant L(i-1)$, hence the second term in the sum above is greater or equal to

$$L(\lceil n/2\rceil - 2) - \sum_{i=1}^{\lceil n/2\rceil-2} \frac{L(\lceil n/2\rceil - 2)}{2^{\lceil n/2\rceil-2-i+1}} = L(\lceil n/2\rceil - 2) \cdot \left(1 - \sum_{i=1}^{\lceil n/2\rceil-2} \frac{1}{2^{\lceil n/2\rceil-1-i}}\right) =$$

$$= L(\lceil n/2\rceil - 2) \cdot \left(1 - \sum_{i=1}^{\lceil n/2\rceil-2} \frac{1}{2^i}\right) \geqslant L(\lceil n/2\rceil - 2) \cdot \left(1 - \sum_{i=1}^{\infty} \frac{1}{2^i}\right) = L(\lceil n/2\rceil - 2) \cdot 0 = 0.$$

Therefore, the total consumption of Alice is at least $\left(x + \frac{x}{2} + \frac{x}{4} + \ldots + \frac{x}{2^{\lceil n/2\rceil-2}}\right)$, which is exactly the same as the left-hand side of (15). Hence, if (15) holds, then Alice cannot do any more attacks. And this means that Bob wins, because having $\lceil n/2\rceil - 1$ segments of length x allows Bob to sustain $\lceil n/2\rceil - 2$ attacks. The proof of the theorem is complete. $\square$

The solution of (15) can be written explicitly as $x = \dfrac{2^{\lceil n/2\rceil-2}}{2^{\lceil n/2\rceil-1} - 1}$. Then the value of a in this strategy is

$$a = \frac{2^{\lceil n/2\rceil-2}}{2^{\lceil n/2\rceil-1} - 1} \cdot (\lceil n/2\rceil - 1) + 1.$$

Note that $x > 1/2$ in all such strategies. Therefore, for $n > 12$ we get $a > 1/2 \cdot 6 + 1 = 4$, which is worse than the generic approach described in theorem 4.2. However, for $n \leqslant 12$ we get non-trivial results shown in the table below.

n	3, 4	5, 6	7, 8	9, 10	11, 12
$a(R_n) \leqslant$	2	$2 + 1/3$	$2 + 5/7$	$3 + 2/15$	$3 + 18/31$

Table 1: Upper bounds on the values of a given by the family of strategies from theorem 7.1

8 Conclusion

In this paper we have established several new results for the tree game connected to the open problem about the gap between the Kolmogorov complexity and a priori probability of a bit string. We have found the exact values of $a(R_3)$ and $a(R_4)$, proved a lower bound on $a(R_5)$ which distinguishes the game on this tree from the game on R_4 and constructed a family of non-trivial strategies for Bob in games with $n \leqslant 12$ which are optimal for $n \leqslant 4$ and therefore may be close to optimal for $5 \leqslant n \leqslant 12$. These results, though modest, provide a new insight on the game and may become a foundation for successful further research in this area.

9 Literature and references

1. Shen, A., Uspensky, V. A., and Vereshchagin, N. (2017). Kolmogorov complexity and algorithmic randomness, volume 220

2. Gacs, P. (1983). On the Relation Between Descriptional Complexity and Algorithmic Probability

3. Day, A. Increasing the gap between descriptional complexity and algorithmic probability. Transactions of the American Mathematical Society, 363(10):5577–5604, 2011