

# Pricing complexity options

Malihe Alikhani<sup>a</sup>, Bjørn Kjos-Hanssen<sup>b,\*</sup>, Amirarsalan Pakravan<sup>c</sup> and Babak Saadat<sup>d</sup>

<sup>a</sup>*Department of Computer Science, Rutgers, The State University of New Jersey, Piscataway, NJ, USA*

<sup>b</sup>*Department of Mathematics, University of Hawai‘i at Mānoa, Honolulu, HI, USA*

<sup>c</sup>*Department of Finance, George Washington University, Washington, DC, USA*

<sup>d</sup>*Kash Co, West Hollywood, CA, USA*

**Abstract.** We consider options that pay the complexity deficiency of a sequence of up and down ticks of a stock upon exercise. We study the price of European and American versions of this option numerically for automatic complexity, and theoretically for Kolmogorov complexity. We also consider run complexity, which is a restricted form of automatic complexity.

Keywords: Automatic complexity, Kolmogorov complexity, options, option pricing

## 1. Introduction

In this article we consider the pricing of American and European options paying the *complexity deficiency*, or intuitively the lack of complexity, of a sequence of up and down ticks for a financial security. The complexity notions we consider are plain and prefix-free Kolmogorov complexity, nondeterministic automatic complexity, and run complexity.

### 1.1. Motivation

We believe it may be of value in finance to have some notions of the complexity of a price path. Agents may want to insure against too complex or too simple price paths for a stock, for example. A very simple or complex path may be a sign that something is going on that the agent is not aware of.

Weather is somewhat periodic, and automatic complexity measures periodicity, to some extent. Hence a complexity option may be used as a weather derivative.

Casino owners may want to ensure that their casinos are truly random, so as to avoid unexpected losses. In general, anyone who makes an assumption of randomness may want to hedge that, as true randomness is not easy to guarantee, or even completely well-defined.

**Automatic complexity: between two extremes.** Of course, we can insure against certain types of non-randomness in simple ways. We can insure against a dramatic fall of a stock price by selling the stock short. This corresponds to *run complexity* (Section 3.2). At the other end, one cannot use *Kolmogorov complexity* (Section 2) as a basis for the security, because Kolmogorov complexity is not computable. The *nondeterministic automatic complexity*, being both

- powerful enough to discern a variety of patterns, and at the same time
- single-exponential time computable,

may be a promising middle ground.

### 1.2. Automatic complexity and the idea of complexity deficiency

Kolmogorov complexity is an important notion that in a way is to complexity as Turing computability is to computability. It is computably approximable, but

---

\*Corresponding author: Bjørn Kjos-Hanssen, Department of Mathematics, University of Hawai‘i at Mānoa, Honolulu, HI 96822, USA. Tel.: +1 808 956 8595; Fax: +1 808 956 9139; E-mail: bjoernkh@hawaii.edu.

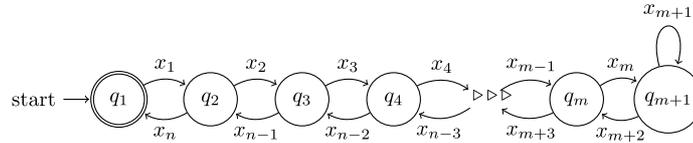


Fig. 1. A nondeterministic finite automata that only accepts one string  $x = x_1x_2x_3x_4 \dots x_n$  of length  $n = 2m + 1$ .

unfortunately not computable. As a remedy, Shallit and Wang (2001) defined the *automatic complexity* of a finite binary string  $x = x_1 \dots x_n$  to be the least number  $A_D(x)$  of states of a deterministic finite automaton  $M$  such that  $x$  is the only string of length  $n$  in the language accepted by  $M$ .

Automatic complexity is computable, but it does have a couple of awkward properties that make us want to tweak its definition. First, many of the automata used to witness the complexity have a *dead state* whose sole purpose is to absorb any irrelevant or unacceptable transitions. Second, some strings  $x = x_1 \dots x_n$  have a different complexity from their reverse  $x_n \dots x_1$ . For instance (Hyde and Kjos-Hanssen 2014; Hyde and Kjos-Hanssen 2015),

$$A_D(011100) = 4 < 5 = A_D(001110).$$

We tweak the definition of automatic complexity by introducing nondeterminism.

**Definition 1.** (Hyde and Kjos-Hanssen (2015)). The nondeterministic automatic complexity  $A_N(w)$  of a word  $w$  is the minimum number of states of an NFA  $M$  (having no  $\epsilon$ -transitions) accepting  $w$  such that there is only one accepting path in  $M$  of length  $|w|$ .

Moreover, and most importantly for the present paper,  $A_N$  gives rise to a striking instance of the idea of *complexity deficiency*:

**Theorem 2.** (Hyde (2013) and Hyde and Kjos-Hanssen (2015)). *The nondeterministic automatic complexity  $A_N(x)$  of a string  $x$  of length  $n$  satisfies*

$$A_N(x) \leq b(n) := \lfloor n/2 \rfloor + 1.$$

**Proof sketch.** The proof is essentially contained in Fig. 1, although we must modify the picture slightly if  $x$  has even length.  $\square$

**Definition 3.** The *nondeterministic automatic complexity deficiency* of a string  $x$  is defined by

$$D_n(x) = b(n) - A_N(x),$$

with  $b(n)$  as in Theorem 2. Sometimes we write  $D(x)$  for  $D_n(x)$ .

Experimentally we have found that about half of all strings have  $D_n(x) = 0$  (Hyde and Kjos-Hanssen, 2015). We call such strings *complex*, and other strings *simple*, herein.

### 1.3. Option types: Perpetual, American, European

We shall consider the following types of options and their prices.

$V$ . This is the price of the perpetual option that pays out the deficiency  $D_n(x)$  when we exercise the option at a time  $n$ . (*Perpetual* here means that we can exercise the option at any time step labeled by a nonnegative integer.) The price of a perpetual option is the supremum, over all exercise policies  $\tau$ , of the expected payoff when using  $\tau$ . There is no restriction that  $\tau$  be computable (and in fact computable before the next market time step occurs), but if that were to become an issue one would presumably change the definition accordingly.

$V_n$ . This is the price of an *American* option that we can exercise at any time step labeled by an integer between 0 and  $n$ .

$W_n$ . This is the price of the *European* option with expiry  $n$ ; in this case we must exercise the option at time  $n$ , if at all, and so  $W_n = \mathbb{E}(\max\{D_n, 0\})$ . We assume the underlying probability distribution is given by the fair-coin measure. In a finance setting it could more generally be given by the risk-neutral measure determined from a stock price process.

We have

$$\mathbb{E}D_n \leq W_n \leq V_n \leq V,$$

and

**Theorem 4.**

$$\sup_n \mathbb{E} D_n \leq \sup_n V_n \leq V \leq \mathbb{E} \sup_n D_n.$$

**Proof.** For the first inequality, it suffices to show

$$\mathbb{E} D_n \leq V$$

for each  $n$ . This holds because one possible exercise policy is the static strategy of exercising at time  $n$  no matter what.

For the third inequality, there are two cases.

Case 1:  $\sup_n D_n$  is almost surely finite. Note that  $D_n$  is integer-valued, so  $\sup_n D_n$  will be realized at some finite stage  $n_0$ . Let us call *magically prescient* the strategy which waits for  $\sup_n D_n$  to be realized and then exercises the option. By contrast, an *exercise policy* should be a stopping time, i.e., it should not depend on future outcomes. We see that the payoff from the magically prescient strategy has a higher price than any exercise policy. It follows that  $V \leq \mathbb{E} \sup_n D_n$  in this case.

Case 2:  $\mathbb{P}(\sup_n D_n = \infty) > 0$ . Then  $\mathbb{E} \sup_n D_n = \infty$  and so we are done.  $\square$

**Remark 5.** In Case of Theorem 4, if  $\mathbb{P}(\sup_n D_n = \infty) = \varepsilon > 0$  then we can even assert that  $V = \infty$ . Indeed if  $V < \infty$  then we can buy the option, and wait for  $D_n > V/\varepsilon + 1$ . The expected payoff is at least

$$(\varepsilon)(V/\varepsilon + 1) = V + \varepsilon > V,$$

which would create an arbitrage.

We shall consider several complexity notions:

- prefix-free Kolmogorov complexity  $K$ ,
- plain Kolmogorov complexity  $C$ , and
- nondeterministic automatic complexity  $A_N$ .

For each notion we first define one or more suitable deficiency notions  $D_n(x)$ : for instance,  $D_n(x) = n + c_C - C(x)$  for a suitable constant  $c_C$  for  $C$ , and  $D_n(x) = \lfloor n/2 \rfloor + 1 - A_N(x)$  for  $A_N$ . The following questions are natural for each of these deficiency notions:

- Does the price of the European option tend to  $\infty$ ?
- Does the price of the American option tend to  $\infty$ ?
- Does the American option for  $C$  have an efficiently computable exercise policy?
- If so, is the increase in value of the American option necessarily slow (say, logarithmic)?

**2. Kolmogorov complexity****2.1. Plain complexity  $C$** 

Let  $c_C$  be the least constant  $c_C$  such that  $C(x \upharpoonright n) \leq n + c_C$  for all strings  $x$  of any length  $n$ . If we define  $D_n(x) = n + c_C - C(x \upharpoonright n)$  for  $x$  of length  $n$ , then  $D_n(x) \geq 0$  for all  $x$ , and  $D_n(x) = 0$  does occur. This is theoretically pleasant. Deficiencies are nonnegative and can be zero. Of course,  $c_C$  depends on the version of the plain length-conditional Kolmogorov complexity  $C(\cdot \upharpoonright \cdot)$  that we use. In this setting, we have

**Theorem 6.**  $\sup_n \mathbb{E} D_n < \infty$ .

**Proof.** Fix  $n$ . For any  $a$ , there are only  $2^a - 1$  binary strings of length at most  $a$ . All descriptions witnessing complexity (given  $n$ ) being at most  $a$  must be among them, so at most  $2^a - 1$  many strings have complexity (given  $n$ ) of at most  $a$ . Applying this to  $a = n + c_C - k$ , at most  $2^{n+c_C-k} - 1$  strings  $x$  (in particular, at most that many strings of length  $n$ ) satisfy  $D_n(x) \geq k$ . That is, by Downey and Hirschfeldt (2010, Corollary 6.1.4),

$$\mathbb{P}(D_n(x) \geq k) < 2^{c_C-k}.$$

Then we have

$$\begin{aligned} \mathbb{E} D_n &= \sum_{k=0}^{\infty} k \mathbb{P}(D_n = k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(D_n \geq k) < \sum_{k=1}^{\infty} 2^{c_C-k} = 2^{c_C}. \quad \square \end{aligned}$$

It turns out that for options expiring at time  $n$ , there is a significantly better exercise policy than the static strategy of waiting until the very end:

**Theorem 7.** For plain Kolmogorov complexity,  $\sup_n V_n = \infty$ , even if we require efficient computation of the exercise policy.

The idea of the proof is to use complexity oscillations, first observed by Martin-Löf (1971): when the initial part of a string  $x$  is a binary encoding of the length of  $x$ , the plain Kolmogorov complexity of  $x$  will be low.

**Proof.** Martin-Löf (1971) showed that deficiency is unbounded for all reals: for each  $X$  and  $b$  there is an  $n$  with  $D(X \upharpoonright n) > b$ . We can computably identify such an  $n$ . The well known idea is that we take a prefix  $X \upharpoonright m$ ; consider it as a binary representation of a

length  $\ell < 2^m$ ; and then consider  $\sigma = X \upharpoonright \ell$ . Since the beginning of  $\sigma$  is known just from the length of  $\sigma$ ,  $\sigma$  is compressible. This translates into an exercise policy for our option: at time  $m$  we decide on the time  $\ell$  at which we are going to exercise. The strategy just described is efficient, since we decide at time  $m \ll \ell$  to exercise at time  $\ell$ .  $\square$

**Remark 8.** Since  $C(x \upharpoonright n) \leq^+ C(x)$ , Theorem 7 holds equally for length-conditional plain Kolmogorov complexity, and Theorem 6 also holds if we consider plain Kolmogorov complexity that is not length-conditional.

## 2.2. Prefix-free complexity $K$ with $C$ -style deficiency

Let  $K$  denote prefix-free Kolmogorov complexity. With  $D_n(x) = n - K(x)$ , there is no limiting deficiency distribution in this case (or one could say the deficiency is in the limit  $-\infty$  almost surely). That is,  $K(w) \geq |w| - c$  for almost all  $w$ , for any  $c$ . Indeed, for each  $c \in \mathbb{Z}$ ,

$$\lim_{n \rightarrow \infty} \frac{|\{\sigma \in 2^n : K(\sigma) \geq n - c\}|}{2^n} = 1,$$

as is easily shown using  $\sum_{\sigma} 2^{-K(\sigma)} < 1$ . If the lim sup of the complement is  $\delta > 0$ , then for each  $\varepsilon > 0$  there exist  $N_k$  with

$$\begin{aligned} 1 &\geq \sum_{\sigma} 2^{-K(\sigma)} = \sum_n \sum_{|\sigma|=n} 2^{-K(\sigma)} \\ &> \sum_k \delta(1 - \varepsilon) 2^{N_k} 2^{-(N_k - c)} \\ &= (1 - \varepsilon) \delta \sum_k 2^c = \infty. \end{aligned}$$

**Theorem 9.** For prefix-free Kolmogorov complexity the price of the perpetual option that pays  $D_n - a$  is at most  $2^{-a}$ .

**Proof.** We have

$$\begin{aligned} \mathbb{P}(\sup_n D_n - a > c) &= \mathbb{P}(\exists n K(X \upharpoonright n) \\ &< n - c - a) \leq 2^{-c-a}. \end{aligned}$$

Let  $D_n^+ = \max\{D_n - a, 0\}$ . Since we would not exercise an option giving negative payoff, it follows that

$$\begin{aligned} V &\leq \mathbb{E}(\sup_n D_n^+) = \sum_{c=0}^{\infty} c \mathbb{P}(\sup_n D_n^+ = c) \\ &= \sum_{c=1}^{\infty} c \mathbb{P}(\sup_n D_n^+ = c) = \sum_{c=1}^{\infty} \mathbb{P}(\sup_n D_n^+ > c) \\ &= \sum_{c=1}^{\infty} \mathbb{P}(\sup_n D_n - a > c) \leq \sum_{c=1}^{\infty} 2^{-c-a} = 2^{-a}. \quad \square \end{aligned}$$

## 2.3. Prefix-free complexity $K$ with its natural notion of deficiency

**Theorem 10.** (Deficiency based on an upper bound for  $K$ ). If we fix a constant  $c_K$  such that for prefix-free Kolmogorov complexity  $K$ ,  $K(x) \leq n + K(n) + c_K$  for all  $x$  of any length  $n$ , and let  $D_n(x) = n + K(n) + c_K - K(x) \geq 0$ , then  $\mathbb{E}D_n$  is bounded but  $V_n \rightarrow \infty$ .

**Proof.** The same proof as for Theorem 6 but using an analogous property shows that  $\mathbb{E}D_n$  is bounded. In this case, however,  $\sup_n D_n(X \upharpoonright n)$  will be  $\infty$  for almost all  $X \in 2^\omega$ . In fact Li and Vitányi showed  $D_n(X \upharpoonright n) > \log n$  for infinitely many  $n$  for almost all  $X$ .

Solovay showed that  $\liminf D_n(X \upharpoonright n)$  will be finite (Miller and Yu, 2011).

$V = \infty$  in this case since we can simply wait for a sufficiently high  $D_n$  value. What about  $V_n$ ? Consider an arbitrary constant, which for expository vividness we will take to equal 17. Almost surely there will be an  $n$  with  $D_n(X \upharpoonright n) \geq 17$ . Therefore for each  $\varepsilon$  there is an  $n_0$  such that

$$\mathbb{P} \bigcup_{n \leq n_0} \{D_n(X \upharpoonright n) \geq 17\} \geq 1 - \varepsilon$$

and so  $V_{n_0} \geq 17(1 - \varepsilon)$ . Moreover  $V_n \leq V_{n+1}$  for American options. So  $V_n \rightarrow \infty$  in this case. The exercise policy would be to wait for  $D_n = 17$  to occur and then exercise.  $\square$

An overview of the deficiency option prices is given in Table 1.

**Remark 11.** Of course, one does not need to only consider deficiencies. One could consider an option paying out  $K(x) - n$ . This value will go to infinity, but how fast? What is our exercise policy if we are not given access to  $K$ ? Another possibility is to consider dips in complexity associated with the Kolmogorov structure function (N. K. Vereshchagin and Vitányi 2004) and its automatic complexity variant (Kjos-Hanssen, 2016).

Table 1

Infinity and finiteness of option prices for various complexity deficiencies  $D_n(x)$ , for strings  $x$  of length  $n$ . The conclusions labeled by  $\therefore$  (“therefore”) follow from the inequalities  $\sup_n \mathbb{E}D_n \leq \sup_n V_n \leq \mathbb{E} \sup_n D_n$  (Theorem 4)

$D_n$	$\sup_n \mathbb{E}D_n$	$\sup_n V_n$	$\mathbb{E} \sup_n D_n$
$n + c_K - K(x)$	$\therefore < \infty$	$\therefore < \infty$	$< \infty$ (Theorem 9)
$n + K(n) + c_K - K(x)$	$< \infty$ (Theorem 10)	$\infty$ (Theorem 10)	$\therefore \infty$
$n + c_C - C(x)$	$< \infty$ (Theorem 6)	$\infty$ (Theorem 7)	$\therefore \infty$
$n + c_C - C(x   n)$	$< \infty$ (Theorem 6)	$\infty$ (Theorem 7)	$\therefore \infty$
$\lceil n/2 \rceil + 1 - A_N(x)$	$< \infty?$ (Conjecture 15)	$\infty?$ (Conjecture 15)	$\therefore \infty?$

2.4. Using runs

**Remark 12.** An anonymous referee suggested the following approach to obtaining results of the form  $V_n \rightarrow \infty$ . Let  $R_n$  be the longest run of 0s in a string of length  $n$  and let  $\mathbb{E}$  and  $\text{Var}$  denote expectation and variance with respect to the uniform distribution on  $\{0, 1\}^n$ . Now, if  $U$  is a universal prefix-free machine, we can define another machine  $M$  by the following algorithm: on input  $x^*$ , simulate  $U$ , and if  $U(x^*) = x$ , then

$$M(x^*) = f(x) := x 0^{\lfloor \log_2 |x| \rfloor - c}$$

for a fixed constant  $c$ . The domain of  $M$  equals the domain of  $U$ , hence  $M$  is also a prefix-free machine. Thus

$$K(x 0^{\lfloor \log_2 |x| \rfloor - c}) \leq^+ K(x).$$

Let now  $m = |f(x)| = |x| + \lfloor \log_2 |x| \rfloor - c$  and  $y = f(x)$ . Since  $K(n) \leq^+ K(m)$  by the choice of  $m$ , we have

$$\begin{aligned} K(y) &\leq n + K(n) + c_K \leq^+ n + K(m) \\ +c_K &= (m + K(m) + c_K) - (m - n) \end{aligned}$$

and

$$C(y) \leq^+ n + c_C = m + c_C - (m - n).$$

Now we employ the trading strategy whereby we wait until our input is of the form  $x 0^{\lfloor \log_2 |x| - c \rfloor}$ , and then exercise. By Theorem ?? below,  $|\mathbb{E}(R_n) - \log_2 n|$  and  $\text{Var}(R_n)$  are both bounded by a constant  $c$ . By the argument in Section 3.2 below, with high probability we will be able to exercise. Thus for American options, with payoff  $D_n(x)$  either  $n + K(n) + c_K - K(x)$  or  $c_C - C(x)$ , we obtain  $V_n \rightarrow \infty$ .

**Theorem 13.** (Boyd (1972)) *Let  $R_n$  be the longest run of heads in a binary sequence of length  $n$  distributed according to the Bernoulli distribution with parameter  $1/2$ . Let  $\log = \ln$ . Then*

$$\begin{aligned} \mathbb{E}(R_n) &= \log_2 n + \frac{\gamma}{\ln 2} - \frac{3}{2} \\ &+ \varepsilon_1(\log n / \log 2) + r_1(n), \end{aligned}$$

where  $\varepsilon_1(\alpha)$  is a function of period 1 which satisfies  $|\varepsilon_1(\alpha)| < 2 \times 10^{-6}$  for all  $\alpha$ , and  $r_1(n) = O(n^{-1}(\log n)^4) \rightarrow 0$ . Moreover,

$$\begin{aligned} \text{Var}(R_n) &= \frac{1}{12} + \frac{\pi^2}{6(\log 2)^2} \\ &+ \varepsilon_2 \left( \frac{\log n}{\log 2} \right) + O(n^{-1}(\log n)^5), \end{aligned}$$

where  $\varepsilon_2(\alpha)$  has period 1, and  $|\varepsilon_2(\alpha)| < 10^{-4}$  for all  $\alpha$ .

3. Computable forms of complexity

3.1. Automatic complexity

Now the goal is to price the European/American option that pays the nondeterministic automatic complexity deficiency  $D_n$  of the movements of a stock from time 0 to the time  $n$  when the option is exercised. We suspect that finding the exact price is a computationally intractable problem, both because of the conjectured intractability of computing automatic complexity (Hyde and Kjos-Hanssen 2015), and because of the exponential number of price paths to consider.

The interest rate  $r$  can be set to 0 or to a positive value. For pedagogical reasons, Shreve (2004) uses  $r = 1/4$  for his main recurring example, and we sometimes adopt that value as well.

- For  $n = 0$  the option would pay 0 as there are no simple strings, and moreover the situation is anyway already known at time 0.
- For  $n = 1$  the actual string (0 or 1) is not known at time 0 but it does not affect the payoff, which is 0 either way as there are no simple strings.

- For  $n = 2$ , with up-factor  $u = 2$ , down-factor  $d = \frac{1}{2}$ , and  $r = 1/4$ , there is a risk-neutral probability of  $1/2$  of one of the strings 00, 11, both of which pay \$1. So the value is

$$(1 + r)^{-2} \cdot \frac{1}{2} \cdot 1 = \frac{16}{50}.$$

In general when the risk-neutral probabilities are  $1/2$  each for up and down, then the value of the option is directly related to the distribution of the deficiency  $D_n$ :

$$\sum_{d=0}^{n/2} d \cdot \mathbb{P}(D_n = d) \cdot (1 + r)^{-n} = \mathbb{E}(D_n) \cdot (1 + r)^{-n}.$$

If  $D_n$  happened to be Poisson for large  $n$ , this is approximately  $\lambda(1 + r)^{-n}$ , which is decreasing in  $n$ . However, we have just seen that the value for  $n = 2$  is higher than for  $n = 0$  and  $n = 1$ .

**Remark 14.** For an American version, one question is whether to exercise the option at time  $n = 2$  after having seen 00. If we exercise we get \$1. Otherwise the deficiency can at most go up by 1 each time step, whereas the interest factor with  $r = 1/4 > 0$  is exponential, so an upper bound for our payoff is

$$(n/2)(1 + r)^{-n} = \frac{n}{2}e^{-n \ln(5/4)}.$$

This expression is maximized at  $n = 4$  and at  $n = 5$ . Both places it takes the value 0.8192.

To obtain a reasonable level of abstraction it is valuable to consider infinite price paths and associate a finite complexity deficiency with them. We can do so if the nondeterministic automatic complexity deficiencies of prefixes of an infinite binary sequence are almost surely bounded (Conjecture 3.1; see also Table 1).

Table 2

Static versus dynamic exercise policies for nondeterministic automatic complexity

Length	$\mathbb{E}D_n$	$\leq$	$V_n$
0	0	=	0
2	0.5	=	0.5
4	0.625	<	0.75
6	0.687	<	0.875
8	0.765	<	1.070
10	0.791	<	1.191
12	0.720	<	1.236

**Conjecture 15.** For nondeterministic automatic complexity  $A_N$ ,

$$\mathbb{P}(\sup_n D_n < \infty) = 1, \text{ and yet } \sup_n V_n = \infty.$$

**Remark 16.** Pakravan and Saadat (2013) studied a perpetual American option that pays the complexity deficiency of the sequence of up and down ticks (considered as 1s and 0s) upon exercise. With interest rate set to zero ( $r = 0$ ) the price of this security may be infinity, based on tentative numerical evidence. That is, for  $A_N$ ,

$$\sup_n V_n = \infty,$$

although  $\mathbb{E}D_n$  seems to approach a finite limit (see Table 2). For positive interest rates the price is finite (see Remark 4). They found numerical evidence that for  $r = 1/4$  the price is 0.47. See Fig. 2 for the deficiencies of strings of length at most 4, and Fig. 3 for corresponding calculated option prices. The price of the American option with expiry  $2k$  and expiry  $2k + 1$  are the same, as is easy to prove.

**Definition 17.** Let  $V^{(n)}$  be the price of the European option paying the nondeterministic automatic complexity deficiency  $D(x)$  for the price path  $x$  of length  $n$ .

*Decision problem:* PRICE.

*Instance:* A pair of nonnegative integers  $n$  and  $k$  with

$$0 \leq \frac{k}{2^n} \leq \lfloor n/2 \rfloor + 1.$$

*Question:* Is  $V^{(n)} \geq k/2^n$ ?

Recall that E is the class of single-exponential time decidable decision problems.

**Theorem 18.** PRICE is in E.

**Proof.** Hyde and Kjos-Hanssen (2015) considered the problem DEFICIENCY of deciding whether, given an integer  $k$  and a sequence  $x$ , the nondeterministic automatic complexity deficiency  $D(x)$  satisfies  $D(x) \geq k$ . They showed that DEFICIENCY is in E. Since there are only single-exponentially many price paths of length  $n$ , the usual backwards recursive algorithm for option pricing in the binomial model (Shreve, 2004) gives the theorem.  $\square$

The same proof shows that the analogous statement to Theorem 18 for American options holds as well.

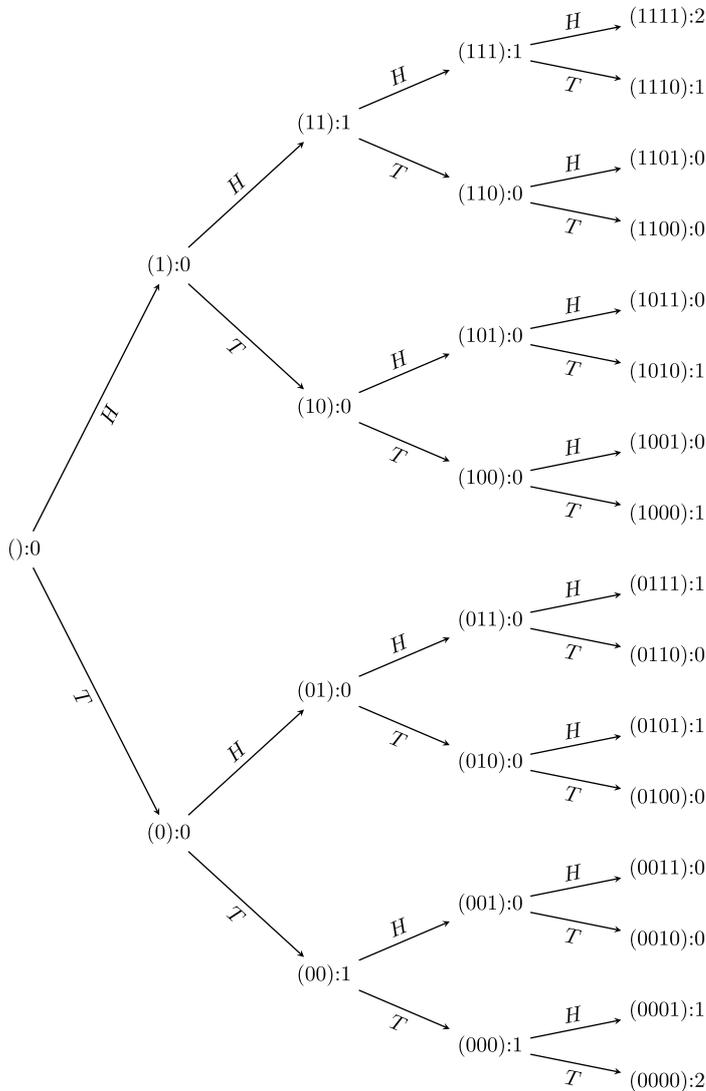


Fig. 2. Deficiency tree for  $n = 4$ , see Remark 16.

### 3.2. Run complexity

If the payoff of our option is just the longest run of heads then Alikhani (2014) showed that the price of the option is  $\Theta(\log_2 n)$ . This corresponds to automata that always proceed to a fresh state, except that one state may be repeated (namely, the state of the longest run).

**Definition 19.** The run complexity  $C_R$  of a binary sequence  $x$  is defined by  $C_R(x) = n + 1 - r$ , where  $n$  is the length of  $x$  and  $r$  is the length of the longest run of 0s or 1s in  $x$ .

This complexity notion has the advantage that it is efficiently computable. Kjos-Hanssen (2014) studied it in more detail and also considered multiple runs, as in the Wald–Wolfowitz runs test.

In the rest of this subsection we give the argument of Alikhani (2014). We assume familiarity with basic discrete options (Shreve, 2004). A coin tossing sequence is  $\omega_1 \dots \omega_N$  where each  $\omega_i \in \{H, T\}$ . (Read  $H$  as “heads” and  $T$  as “tails”.)

**Definition 20.** For each  $0 \leq n \leq N$ , the current run of heads in the coin tossing sequence  $\omega_1 \dots \omega_n$  is defined by

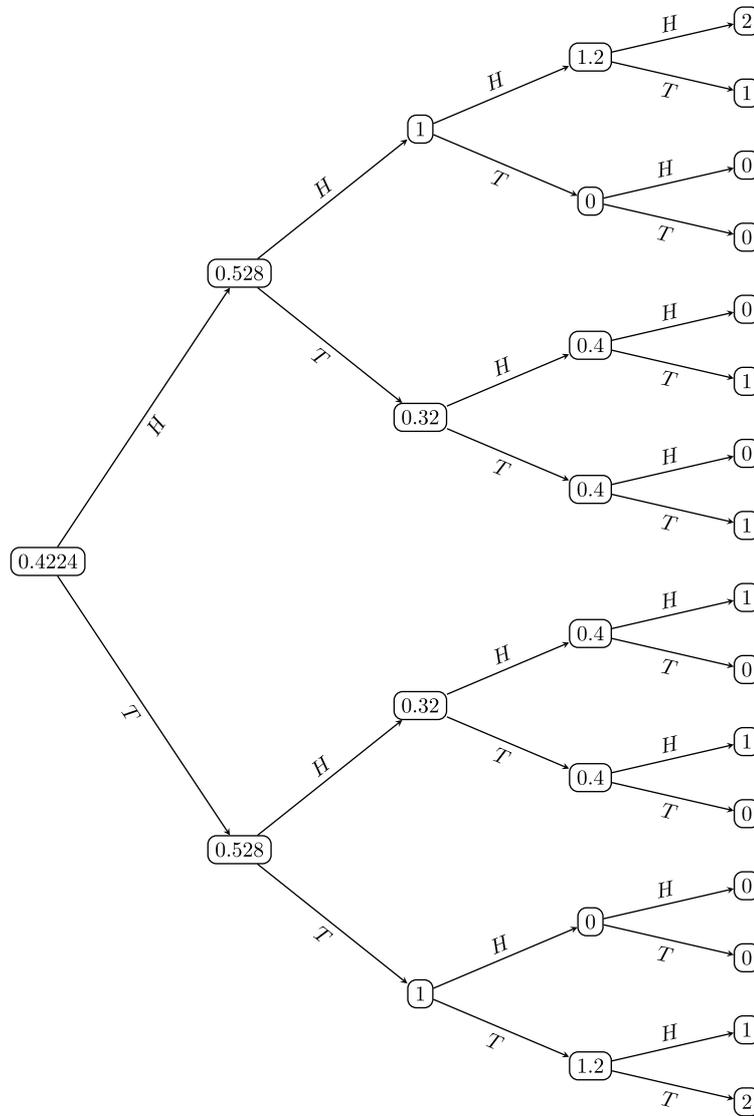


Fig. 3. Option prices corresponding to Fig. 2.

$$G_n(\omega) = \max\{r : \omega_{n-r+1} = \dots = \omega_n = H\}.$$

The *run option* is the American option where the payoff when exercised at time  $n \leq N$  is  $G_n(\omega)$ . Let  $V^A$  be the price of the run option. Define a stopping time  $\tau_t$  by

$$\tau_t(\omega_1 \dots \omega_N) = \min\{s : G_s = [\mathbb{E}(R_N)] - t\},$$

where  $R_N$  is the longest run of heads in a coin tossing sequence of length  $N$ .

Thus, the trading strategy corresponding to  $\tau_t$  is to wait for a run of heads that is almost as long as we

ever expect to see before time  $N$ , with “almost” being qualified and measured by the parameter  $t$ .

**Definition 21.** Let  $[x]$  denote the nearest integer of  $x$ . In particular,  $[x]$  is an integer  $k$  with  $k - 1 \leq x \leq k + 1$ .

**Theorem 22.** Given  $N$  there is a deterministic choice of  $t = t_N$  such that there is a sequence of numbers  $\varepsilon_N$  with  $\lim_{N \rightarrow +\infty} \varepsilon_N = 0$ , and constants  $c_2$  and  $c$ , such that for large  $N$ ,

$$\mathbb{E}(G_{\tau_N}) \geq (\log_2 N - c_2 - c\sqrt[3]{\ln N})(1 - \varepsilon_N).$$

**Proof.** The price process  $V_n^A$  for the run option satisfies the American risk-neutral formula

$$V_n^A = \max_{\tau \in \mathcal{S}_n} \mathbb{E}_n[\mathbb{I}_{\tau \leq N} G_\tau], \quad \text{for } n = 0, 1, \dots, N.$$

So for each  $t$ ,

$$V_n^A \geq \mathbb{E}_n[(\mathbb{I}_{\tau_t \leq N}) G_{\tau_t}].$$

Now we find a lower bound on  $\mathbb{E}(G_{\tau_t})$ .

$$\begin{aligned} \mathbb{E}(G_{\tau_t}) &= \mathbb{E}(G_{\tau_t} | G_{\tau_t} > 0) \Pr(G_{\tau_t} > 0) \\ &= ([\mathbb{E}(R_N)] - t) \Pr\{R_N \geq [\mathbb{E}(R_N)] - t\} \\ &\geq ([\mathbb{E}(R_N)] - t) \Pr\{R_N \geq \mathbb{E}(R_N) - t + 1\} \\ &\geq ([\mathbb{E}(R_N)] - t) \Pr\{|R_N - \mathbb{E}(R_N)| \leq (t - 1)\} \\ &\geq ([\mathbb{E}(R_N)] - t) \left(1 - \frac{\text{Var}(R_N)}{(t - 1)^2}\right) \\ &\quad \text{(by Chebyshev's Inequality)} \\ &\geq (\mathbb{E}(R_N) - 1 - t) \left(1 - \frac{\text{Var}(R_N)}{(t - 1)^2}\right). \end{aligned}$$

By Theorem 13,

$$\text{Var}(R_N) = \pi^2/6 \ln^2(2) + 1/12 + r_2(N) + \varepsilon_2(N) \leq 4$$

for large  $N$ . Let  $\mathbb{E}(R_N) = a$ ; then we get

$$\mathbb{E}(G_{\tau_t}) \geq (a - t - 1) \left(1 - \frac{4}{(t - 1)^2}\right). \quad (*)$$

Now we find the  $t = t_N$  such that the right-hand side of (\*) is maximized. The corresponding third degree polynomial has negative discriminant. Therefore it has one real root, which was calculated by Mathematica:

$$t = \frac{\left(\frac{2}{3}\right)^{2/3} \sqrt[3]{9 \ln^2(2) \ln(N) + \sqrt{3} \sqrt{27 \ln^4(2) \ln^2(N) + 4 \ln^6(2)}}}{\ln(2)} - \frac{2 \sqrt[3]{\frac{2}{3}} \ln(2)}{\sqrt[3]{9 \ln^2(2) \ln(N) + \sqrt{3} \sqrt{27 \ln^4(2) \ln^2(N) + 4 \ln^6(2)}}.$$

By the second derivative test, since

$$\frac{d^2}{dt^2}(a - t)(1 - 4/t^2) = -3t^2 - 4 \leq 0,$$

we see that  $t$  maximizes the right-hand side of (\*). We have

$$\lim_{n \rightarrow \infty} - \frac{2 \sqrt[3]{\frac{2}{3}} \ln(2)}{\sqrt[3]{9 \ln^2(2) \ln(n) + \sqrt{3} \sqrt{27 \ln^4(2) \ln^2(N) + 4 \ln^6(2)}} = 0$$

and hence

$$\frac{t}{(4/\ln 2)^{1/3} (\sqrt[3]{\ln N})} \rightarrow 1.$$

Therefore,  $t = t_N \in \Theta(\sqrt[3]{\ln N})$  and so

$$\mathbb{E}(G_{\tau_{t_N}}) \geq (\log_2 N - c_2 - c \sqrt[3]{\ln N})(1 - \varepsilon_N).$$

□

**Corollary 23.**  $V^A \sim \log_2 N$ .

**Proof.**  $V^A$  is bounded below by the expected payoff of the strategy that waits for  $[E(R_N)] - t_N$  heads, with  $t_N$  as in Theorem 22, and then exercises. On the other hand,  $V^A$  is bounded above  $\mathbb{E}(R_N)$ . Therefore

$$\mathbb{E}(R_N) - t_N \leq V^A \leq \mathbb{E}(R_N).$$

By Theorem 13,

$$\mathbb{E}(R_N) = \log_2(N/2) + \gamma/\ln 2 - 1/2 = \log_2 N + O(1),$$

and so by Theorem 22,

$$\log_2 N - c_2 - \sqrt[3]{\ln N} \leq V^A \leq \log_2 N + O(1).$$

Dividing by  $\log_2 N$  we get

$$1 - o(1) \leq \frac{V^A}{\log_2 N} \leq 1 + o(1).$$

□

#### 4. Robustness

We now consider whether, in the phrase of an anonymous referee,

small perturbations on input sequences can have drastic effects on our studied measurements of complexity.

In other words, whether errors in the measurement of a sequence will lead to large errors in the calculated complexity. Let  $d(x, y)$  denote the Hamming distance between two sequences of the same length  $x$  and  $y$ .

**Run complexity.** Here a change in a single bit sometimes cuts the longest run in half. That is, if  $d(x, y) = 1$  then  $C_R(x) = n - r_x$  and  $C_R(y) = n - r_y$  where  $r_x \leq 2r_y + 1$ .

On the other hand, since the longest run will only be about  $\log_2 n$  (Boyd, 1972), a random change in a single random bit will tend to leave the complexity unchanged.

**Automatic complexity.** Here we have numerical evidence that a change in a single bit sometimes has

Table 3

Nondeterministic automatic complexity in the Hamming ball of radius 1 around  $0^n$ ,  $n = 23$

$w$	$A_N(w)$
$0^{23}$	1
$0^{22}1$	2
$0^{21}10$	3
$0^{20}10^2$	4
$0^{19}10^3$	5
$0^{18}10^4$	6
$0^{17}10^5$	7
$0^{16}10^6$	8
$0^{15}10^7$	9
$0^{14}10^8$	8
$0^{13}10^9$	8
$0^{12}10^{10}$	8
$0^{11}10^{11}$	7

large effects. For instance, consider the string  $0^n$  which becomes  $0^a 10^{n-a-1}$ ; see Table 3.

**Kolmogorov complexity.** A change in a single bit will affect the complexity only logarithmically (by at most about  $2 \log n$ ) since a description of the sequence can include hard-coded information about where the changed bit is. Fortnow, Lee, and N. Vereshchagin (2006) studied Kolmogorov complexity with error in detail.

## 5. Enhanced content

**AutoComplex.** This app for Android devices (Bjørn Kjos-Hanssen 2013) lets you look up nondeterministic automatic complexity values of particular strings. The app tells you the complexity of a given string and also provides a “proof” or “witness”.

This witness is a uniquely accepting state sequence, i.e., a sequence of states visited during a run of a witnessing automaton. It is analogous to a shortest description  $x^*$  of a string  $x$ , familiar from the study of Kolmogorov complexity.

The app also provides some extensions of the string suggested by the familiar autocompletion feature used in search engines.

**The Complexity Guessing Game and the Complexity Option Game.** These two online games (Bjørn Kjos-Hanssen 2015a, 2015b) invite the player to guess complexities, or implement an exercise policy for a complexity-based financial option, respectively. The games include graphical displays of millions of the relevant automata.

## Acknowledgments

This work was partially supported by a grant from the Simons Foundation (#315188 to Bjørn Kjos-Hanssen).

This material is based upon work supported by the National Science Foundation under Grant No. 0901020.

## References

- Alikhani, Malihe, 2014. “American option pricing and optimal stopping for success runs”. Project for master of arts in mathematics. 2014.
- Bjørn, Kjos-Hanssen, 2013. “AutoComplex”. <https://play.google.com/store/apps/details?id=edu.hawaii.math.bjoern.auto-complex>. Aug. 10, 2013.
- 2015a. “The Complexity Guessing Game”. <http://math.hawaii.edu/wordpress/bjoern/software/web/complexity-guessing-game/>. Apr. 27, 2015.
- 2015b. “The Complexity Option Game”. <http://math.hawaii.edu/wordpress/bjoern/software/web/complexity-option-game/>. Apr. 27, 2015.
- Boyd, D.W., 1972. “Losing runs in Bernoulli trials”. <http://www.math.ubc.ca/~boyd/bern.runs/bernoulli.html>. 1972.
- Downey, Rodney, G., Denis, R., Hirschfeldt, 2010. Algorithmic randomness and complexity. Theory and Applications of Computability. New York: Springer, 2010, pp. xxviii+855. ISBN:978-0-387-95567-4. DOI:10.1007/978-0-387-68441-3. URL:<http://dx.doi.org/10.1007/978-0-387-68441-3.18>
- Fortnow, Lance, Troy Lee, Nikolai Vereshchagin, 2006. “Kolmogorov complexity with error”. In: STACS 2006. Vol. 3884. Lecture Notes in Comput. Sci. Springer, Berlin, 2006, pp. 137–148. DOI: 10.1007/11672142\_10. URL:[http://dx.doi.org/10.1007/11672142\\_10](http://dx.doi.org/10.1007/11672142_10).
- Hyde, Kayleigh, 2013. “Nondeterministic finite state complexity”. Project for master of arts in mathematics. 2013.
- Hyde, Kayleigh, Bjørn, Kjos-Hanssen, 2014. “Nondeterministic automatic complexity of almost square-free and strongly cube-free words”. In: COCOON 2014. Vol. 8591. Lecture Notes in Comput. Sci. Springer, Heidelberg, 2014, pp. 61–70.
- Hyde, Kayleigh, Bjørn, Kjos-Hanssen, 2015. “Nondeterministic Automatic Complexity of Overlap-Free and Almost Square-Free Words”. In: Electronic Journal of Combinatorics 2015. To appear.
- Kjos-Hanssen, Bjørn, 2014. “Kolmogorov structure functions for automatic complexity in computational statistics”. In: The 8th International Conference on Combinatorial Optimization (COCO 2014). Vol. 8881. Lecture Notes in Comput. Sci. Springer, Berlin, 2014, pp. 652–665.
- Kjos-Hanssen, Bjørn, 2016. “Kolmogorov structure functions for automatic complexity”. In: Theoretical Computer Science 2016. To appear.

- Martin-Löf, Per 1971. "Complexity oscillations in infinite binary sequences". In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 19 1971, pp. 225–230.
- Miller, Joseph, S., Liang, Yu, 2011. "Oscillation in the initial segment complexity of random reals". In: *Adv. Math.* 226.6 2011, pp. 4816–4840. ISSN:0001-8708. DOI:10.1016/j.aim.2010.12.022. URL:<http://dx.doi.org/10.1016/j.aim.2010.12.022>.
- Pakravan, Amirarsalan, Babak, Saadat, 2013. "Nondeterministic finite state complexity and option pricing". Project for master of science in financial engineering. 2013.
- Shallit, Jeffrey, Ming-Wei, Wang, 2001. "Automatic complexity of strings". In: *J Autom Lang Comb.* 6.4 2001. 2nd Workshop on
- descriptive complexity of automata, Grammars and related structures (London, ON, 2000), pp.537–554. ISSN: 1430-189X.
- Shreve, Steven, E., 2004. *Stochastic Calculus for Finance. I.* Springer Finance. The binomial asset pricing model. Springer-Verlag, New York, 2004, pp. xvi+187. ISBN: 0-387-40100-8.
- Vereshchagin, Nikolai, K., Paul, M.B., Vitányi, 2004. "Kolmogorov's structure functions and model selection". In: *IEEE Trans. Inform. Theory* 50.12 2004, pp.3265–3290. ISSN: 0018-9448. DOI:10.1109/TIT.2004.838346. URL:<http://dx.doi.org/10.1109/TIT.2004.838346>.