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

Факультет компьютерных наук
Основная образовательная программа
Прикладная математика и информатика

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
на тему

Случайный шум и изменение колмогоровской сложности

Выполнил студент группы БПМИ141, 4 курса,
Пособин Глеб Игоревич

Научный руководитель:
К.ф.-м.н., старший научный сотрудник,
Шень Александр

Москва 2018

Abstract

Consider a binary string x of length n whose Kolmogorov complexity equals αn for some $\alpha < 1$. We want to increase the complexity of x by changing a small fraction of bits in x. This is always possible: Buhrman, Fortnow, Newman and Vereshchagin showed [Buhrman et al., 2005] that the increase can be at least δn for large n (where δ is some positive number that depends on α and the allowed fraction of changed bits).

We consider a related question: what happens with the complexity of x when we randomly change a small fraction of the bits (changing each bit independently with some probability p)? It turns out that a linear increase in complexity happens with high probability, but the guaranteed increase is smaller than in the case of arbitrary change. We note that the amount of the increase depends on x (strings of the same complexity could behave differently), and give an exact lower and upper bounds for this increase (with $o(n)$ precision).

The proof uses the combinatorial technique that goes back to Ahlswede, Gács and Körner [Ahlswede et al., 1976]. For the reader's convenience (and also because we need a slightly stronger statement) we provide a simplified exposition of this technique, so the paper is self-contained.

Аннотация

Рассмотрим бинарную строку x длины n с колмогоровской сложностью αn для некоторого $\alpha < 1$. Мы хотим увеличить сложность x изменив малую долю битов. Это всегда возможно: Бурман, Фортноу, Ньюман и Верещагин показали [Buhrman et al., 2005], что можно гарантировать увеличение сложности по крайней мере на δn для достаточно больших δ (где δ — некоторое положительное число, которое зависит от α и доли разрешённых изменений).

Мы рассматриваем связанный вопрос: что происходит со сложностью строки x когда мы случайно изменяем небольшую долю битов (то есть, изменяем каждый бит независимо с некоторой вероятностью p)? Оказывается, линейное увеличение сложности происходит с большой вероятностью, но это изменение меньше, чем в случае произвольных изменений. Величина изменения сложности зависит от строки x (строки с одинаковой сложностью могут вести себя по-разному), и даём точные верхние и нижние оценки на изменение сложности с точностью $o(n)$.

Доказательство использует комбинаторную рассуждения, восходящую к Алсведе, Гачу и Кёрнеру [Ahlswede et al., 1976]. Для удобства читателя (а также потому что нам нужно немного более сильное утверждение) мы приводим более простые доказательства нужных результатов, так что работа самодостаточна.

Contents

Changing Kolmogorov complexity with random noise

Gleb Posobin

May 21, 2018

Abstract

Consider a binary string x of length n whose Kolmogorov complexity equals αn for some $\alpha < 1$. We want to increase the complexity of x by changing a small fraction of bits in x. This is always possible: Buhrman, Fortnow, Newman and Vereshchagin showed [Buhrman et al., 2005] that the increase can be at least δn for large n (where δ is some positive number that depends on α and the allowed fraction of changed bits).

We consider a related question: what happens with the complexity of x when we *randomly* change a small fraction of the bits (changing each bit independently with some probability p)? It turns out that a linear increase in complexity happens with high probability, but the guaranteed increase is smaller than in the case of arbitrary change. We note that the amount of the increase depends on x (strings of the same complexity could behave differently), and give an exact lower and upper bounds for this increase (with $o(n)$ precision).

The proof uses the combinatorial technique that goes back to Ahlswede, Gács and Körner [Ahlswede et al., 1976]. For the reader's convenience (and also because we need a slightly stronger statement) we provide a simplified exposition of this technique, so the paper is self-contained.

1 Introduction

The Kolmogorov complexity $C(x)$ of a binary string x is defined as the minimal length of the program that generates x, assuming that we use an optimal programming language that makes the complexity function minimal up to an $O(1)$ additive term (see [Li and Vitányi, 2008, Shen et al., 2017] for details). There are several versions of complexity; we consider the original version, called *plain* complexity. In fact, for our considerations the difference between different versions does not matter, since they differ only by $O(\log n)$ term for n-bit strings, but we restrict ourselves to plain complexity for simplicity.

The complexity of n-bit strings is between 0 and n (we omit $O(1)$ additive terms); most strings have complexity close to n. Consider a string x of length n that has some intermediate complexity, say $0.5n$. Let us change about 1% of bits in x, changing each bit independently with probability 0.01. Does this change increase the complexity of x? It may depend on the bits we change, but it turns out that *for a random change the complexity of the resulting string increases with high probability*: we get a string of complexity at least $0.501n$ with probability at least 99%, for all large enough n (the result is necessarily asymptopic, since Kolmogorov complexity function is defined up to $O(1)$ terms).

This is the type of statement we are interested in; of course, the parameters above are chosen as an example, and the following general statement is true.

Theorem 1. *There exists a strictly positive function $\delta(\alpha, \tau)$ defined for $\alpha, \tau \in (0, 1)$ with the following property: for all sufficiently large n, for every $\alpha \in (0, 1)$, for every $\tau \in (0, 1)$, for $\beta = \alpha + \delta(\alpha, \tau)$, and for every x such that $C(x) \geqslant \alpha n$, the probability of the event*

$$C(N_\tau(x)) > \beta n$$

where $N_\tau(x)$ is a random variable obtained if we change every bit in x independently with probability τ, is at least $1 - 1/n$.

Remark 1. We use the inequality $C(x) \geqslant \alpha n$ (and not an equality $C(x) = \alpha n$) to avoid technical problems: the complexity $C(x)$ is an integer, and αn may not be an integer.

Remark 2. One may consider only $\tau \leqslant 1/2$ since reversing all bits does not change Kolmogorov complexity. For $\tau = 1/2$ the variable $N_\tau(x)$ is uniformly distributed in the Boolean cube $\mathbb{B}^n$, so its complexity is close to n, and the statement is easy (for arbitrary $\beta < 1$).

Remark 3. We use α, τ as parameters while fixing the probability bound as $1 - 1/n$. As we will see, the choice of this bound is not important: we could use a stronger bound (exponentially close to 1) as well.

Now the natural question arises: for which functions $\delta(\alpha, \tau)$ the statement of Theorem 1 holds. In other words, fix α and τ. Theorem 1 guarantees that there exists some $\beta > \alpha$ such that every string x of complexity at least αn is guaranteed to have complexity at least βn after τ-noise (with high probability). *What is the maximal value of β for which such a statement is true?*

Before answering this question, we should note that the guaranteed complexity increase depends on x: for different strings of the same complexity the typical complexity of $N_\tau(x)$ could be different. Here are two opposite examples (with minimal and maximal increase, as we will show).

Example 1. Consider some $p \in (0, 1)$ and Bernoulli distribution B_p on the Boolean cube $\mathbb{B}^n$ (bits are independent; every bit equals 1 with probability p). With high probability the complexity of a B_p-random string is $o(n)$-close to $nH(p)$ (see [Shen et al., 2017, chapter 7]), where $H(p)$ is the Shannon entropy function

$$H(p) = -p \log p - (1 - p) \log(1 - p).$$

After applying τ-noise the distribution B_p is tranformed into $B_{N(\tau,p)}$, where

$$N(\tau, p) = p(1 - \tau) + (1 - p)\tau = p + \tau - 2p\tau$$

is the probability to change the bit if we first change it with probability p and then (independently) change it with probability τ, and the complexity of $N_\tau(x)$ is close (with high probability) to $H(N(\tau, p))$ if the B_p-random string x and the τ-noise are chosen independently. So in this case we have (with high probability) the complexity increase

$$H(p)n \to H(N(\tau, p))n.$$

Note that $N(\tau, p)$ is closer to $1/2$ than p, and H is strictly increasing on $[0, 1/2]$, so indeed some increase happens.

Example 2. Now consider an error-corrrecting code that has $2^{\alpha n}$ codewords and corrects τn errors (this means that the Hamming distance between codewords is greater than $2\tau n$). Such a code may exist or not depending on the choice of α and τ. The basic result in coding theory, Gilbert's bound, guarantees that it exists if α and τ are not too large. Consider some pair of α and τ for which such a code exist; moreover, let us assume that it corrects up to $\tau' n$ errors for some $\tau' > \tau$. We assume also that the code itself (the list of codewords) has small complexity, say, $O(\log n)$. This can be achieved by choosing the first (in some ordering) code with required parameters.

Now take a random codeword of this code; most of the codewords have complexity close to αn. If we randomly change each bit with probability p, then with high probability we get at most $\tau' n$ errors, so decoding is possible and the pair (x, noise) can be reconstructed from $N_\tau(x)$, the noisy version of x. Then the complexity of $N_\tau(x)$ is close to the complexity of the pair (x, noise), which (due to independence) is close to $\alpha n + H(\tau)n$ with high probability. So in this case we have the complexity increase

$$\alpha n \to (\alpha + H(\tau))n.$$

Remark 4. Note that this increase is maximal possible not only for a random independent noise but for any change in x that changes τ-fraction of bits. See below the discussion of the difference between random change and arbitrary change.

Now we formulate the result we promised. It says that the complexity increase observed in Example 1 is the minimal possible: such an increase is guaranteed for every string of given complexity.

Theorem 2. *Let $\alpha = H(p)$ for some $p \in (0, 1)$. Let τ be an arbitrary number in $(0, 1)$. Let β be some number less than $H(N(p, \tau))$. Then for sufficiently large n the following is true: for every string x with $C(x) \geqslant \alpha n$ the string $N_\tau(x)$, obtained from x by changing each bit independently with probability τ, has complexity greater than βN, with probability at least $1 - 1/n$.*

The Example 1 shows that the bound for β in this theorem is optimal. However, the result does not say anything about the boundary case when β is exactly $H(N(p, \tau))$, for this our tools are not precise enough.

Theorem 2 is the main result of the paper. It is proven, as it often happens with results about Kolmogorov complexity, by looking at its combinatorial counterparts. We explain the reduction to a combinatorial statement in the next section. Then in Section 3 we prove the combinatorial statement. In fact, if we are interested in some complexity increase (Theorem 1) a simple argument (suggested by Fedor Nazarov) that uses Fourier transform is enough. A stronger result can be obtained by hypercontractivity techniques. However, for the optimal bound (Theorem 2) we need to use more advanced tools from [Ahlswede et al., 1976] paper.

2 Reduction to combinatorial statements

We want to estimate the complexity increase caused by random noise. Let us consider first a different question. What if we are allowed to change arbitrary bits (only the number of changed bits in bounded) and want to increase complexity? This question was considered in [Buhrman et al., 2005]. It turns out that it is equivalent to a combinatorial statement, and similar equivalence can be used for the case of random change. Let us look first at the case of arbitrary change.

2.1 Arbitrary change

Fix some α and τ. Our goal is to show, for some $\beta > \alpha$, that every string of complexity at least αn can be changed in at most τn position to get some string of complexity at least βn.

Let us reformulate this goal using contraposition: if a string cannot be changed in at most τn positions to get a string of complexity at least βn, then its complexity is at most αn. In other words, consider the set B of all n-bit strings that have complexity less than βn. We want to show that if a Hamming ball of radius τn is contained entirely in B, then the center of this ball has complexity less than αn.

This would imply that the "τn-interior of B", the set A of the centers of τn-balls contained entirely in B, has less than $2^{\alpha n}$ elements. On the other hand, this cardinality bound is enough, the connection works in the reverse direction, too, if we measure complexity with $o(n)$-precision. Indeed, knowing n, βn and τn (their integer parts), we may enumerate the set B; when an entire ball of radius τn is covered by the points of B that are already enumerated, we know that the center of this ball is in A. So A can be enumerated, and each of its elements can be described by specifying (in addition to n and $\lfloor \beta n \rfloor$, which require only $O(\log n)$ bits) its ordinal number in the enumeration, and this number has $\log \# A$ bits. So the $2^{\alpha n}$ bound for the size of A implies $\alpha n + O(\log n)$ bound for the complexity of all elements in A, and the $O(\log n)$ term is negligible since we consider the question with $o(n)$-precision.

Now we can state the combinatorial counterpart of the result from [Buhrman et al., 2005], and the required relation between α and β.

Proposition 1. *Let p be some number in $(0, 1)$ and let $\alpha = H(p)$. Let τ be some positive number so that $p + \tau \leqslant 1/2$, and let $\beta = H(p + \tau)$. Let B be an arbitrary subset of $\mathbb{B}^n$ of size at most $2^{\beta n}$. Let A be a subset of $\mathbb{B}^n$, and for every $x \in A$ the Hamming ball of radius τn with center x is contained in B. Then the cardinality of A does not exceed* $\operatorname{poly}(n) 2^{\alpha n}$.

This is a direct consequence of the Harper's theorem (see, e.g.,[Frankl and Füredi, 1981]) that says that for a set of a given size its c-interior (for some fixed c) is maximal when the set is a Hamming ball (formally speaking, is between two Hamming balls of sizes k and $k+1$ for some k). Or, in dual terms, that the c-neighborhood of a set of a given size is minimal if a set is a Hamming ball. The relation between $2^{\alpha n}$ and $2^{\beta n}$ in the proposition is just the relation between the sizes of balls whose radii differ by τn (we ignore polynomial in n factors for simplicity). Note that $p + \tau \leqslant 1/2$ is needed since otherwise the radius exceeds $n/2$ and then the log-size of the ball is close to n and not $H((p + \tau)n)$. The $\operatorname{poly}(n)$ factor is needed due to the polynomial factor in the estimates of the ball size in terms of Shannon entropy (the ball of radius γn has size $2^{H(\gamma)n}$ up to polynomial in n factors).

We do not go into details here (and do not reproduce the proof of Harper's theorem) since we need this result only to motivate the corresponding relation between combinatorial and complexity statements for a random change.

2.2 Random change

For the random change the corresponding combinatorial statement needs to be changed. Before, for some set B and some number τ, we considered the τn-interior of B, i.e., the points that are contained in B together with its τn-neighborhood. Now we need to consider a bigger set of all points x such that $N_\tau(x)$ gets into B with probability at least $1/n$.[1] Let us state this reduction explicitly.

[1] One could also replace the τ-noise by changing a randomly chosen set of at most τn bits (all sets of size at most τn are considered as equiprobable). Then we have to consider balls where at least $1/n$-th part of the ball is covered by B. This is closer to the case of arbitrary change. Technically this setting is less convenient, but one can translate our results to this language without much difficulties and get the same relation between α, β and τ.

Proposition 2. *Let α and β be some numbers in $(0,1)$, and $\alpha < \beta$. Let $\tau \in (0,1/2)$ be some number. Assume that the following combinatorial statement is true: for every sufficiently large n, for every set $B \subset \mathbb{B}^n$ such that $\#B < 2^{\beta n}$, the set A of all points $x \in \mathbb{B}^n$ that get into B with probability greater than $1/n$ after applying τ-noise, has size $\#A < 2^{\alpha n}$.*

Then, for every $\alpha' > \alpha$, the following complexity statement is true: for all sufficiently large n, for every n-bit string x such that $C(x) \geqslant \alpha' n$, the probability of the event "τ-noise transforms x into a string $y = N_\tau(x)$ such that $C(y) > \beta n$" is at least $1 - 1/n$.

Remark 5. This theorem shows, informally speaking, that the region of (α, β, τ) where the combinatorial statement is true, is contained (may be, except the boundary points, since we require $\alpha' > \alpha$) in the region where the complexity statement is true. In fact, these regions coincide (save the boundary points, maybe), since Example 1 and Theorem 2 provide matching upper and lower bounds.

Proof of Proposition 2. Fix α, β, τ, and α' satisfying the conditions of the proposition. Let n be a sufficiently large integer. Consider the set B of n-bit strings y such that $C(y) < \beta n$. This set contains at most $2^{\beta n}$ elements. Then consider the set A constructed as in the combinatorial statement, i.e., the set of all x such that the probability of the event "$N_\tau(x) \in B$" exceeds $1/n$. The combinatorial assumption guarantees that the size of A does not exceed $2^{\alpha n}$. The set B can be enumerated if n and βn is known. Enumerating B, we get (for each x) increasing lower bounds for the probability of the event $N_\tau(x) \in B$ and can enumerate strings x where this probability exceeds $1/n$. The ordinal number in the enumeration has αn bits (due to the bound for $\#A$), and together with additional $O(\log n)$ needed to specify parameters we get $\alpha n + O(\log n) < \alpha' n$: complexity of each element of A is less then $\alpha' n$. This is exactly the statement we need to prove.[2] □

3 Combinatorial proof

In this section we provide the combinatorial bound, as discussed in the previous section:

Proposition 3. *Let $\alpha = H(p)$ for some $p \in (0,1)$. Let τ be an arbitrary number in $(0,1)$. Let $\beta = H(N(p,\tau))$. Then for every set B of size at most $2^{\beta n}$, the set A of all $x \in \mathbb{B}$ such that the probability of the event "$N_\tau(x) \in B$" is greater than $1/n$, where $N_\tau(x)$ is obtained from x by changing each bit independently with probability τ, has cardinality at most* $\operatorname{poly}(n)2^{\alpha n}$.

Together with the reduction described in the previous section (Proposition 2) this finishes the proof of Theorem 2. (Note that the polynomial factor in Proposition 3 can be absorbed for large n by a small change in α or β, so it does not matter.)

3.1 An easy proof of some decrease

We start with a proof (suggested to us by Fedor Nazarov, personal communication) of a weak version of Proposition 3 showing that for every τ and every $\beta < 1$ there exists some $\alpha < \beta$ such that required bound $\#A \leqslant 2^{\alpha n}$ is valid of every B of size $2^{\beta n}$.

[2] A pedantic reader would say that τ and β may not be computable, and in this case we do not have an algorithm to enumerate B. However, for β we need to known only the maximal integer that is smaller than βn, and this is $O(\log n)$ bits. For τ the situation is more complicated, since the computation of probabilities uses τ. To deal with the problem for non-computable τ, we have to change the bound $1/n$ to something smaller, like $1/2n$, so the proposition should be corrected for noncomputable τ. Still we ignore this problem, since the change of the bound in the combinatorial statement is not important; we will see that the bound can be easily replaced by any bound of the form $1/\operatorname{poly}(n)$.

Every real-valued function on the Boolean hypercube $\mathbb{B}^n$, identified with $\{-1,1\}^n$ and considered as a multiplicative group in this section, can be written in the standard Fourier basis:

$$f(x) = \sum_{S\subset[n]} \widehat{f}_S \chi_S(x),$$

where $\widehat{f}_S$ are Fourier coefficients, $\chi_S(x) = \prod_{i\in S} x_i$. Functions χ_S are characters of the Boolean cube as a multiplicative group. They form an orthonormal basis in the space of real-valued functions on $\mathbb{B}^n$ with respect to the following inner product:

$$\langle f,g\rangle = \frac{1}{2^n}\sum_{x\in\mathbb{B}^n} f(x)g(x) = \mathop{\mathbb{E}}_{x\in\mathbb{B}^n} f(x)g(x)$$

This representation will be useful for us, since the Fourier representation of the convolution of two functions is simply the point-wise product of their Fourier representations: $\widehat{f*g}_S = \widehat{f}_S\,\widehat{g}_S$, where the convolution is defined as

$$(f*g)(x) = \mathop{\mathbb{E}}_{t\in\mathbb{B}^n} f(xt)g(t^{-1}).$$

For a set $B\subset\mathbb{B}^n$ we are interested in the probability

$$N_\tau^B(x) = \Pr[N_\tau(x)\in B].$$

This function is a convolution of the indicator function $\mathbb{1}_B$ of the set B (that is 1 inside the set and 0 outside) and the distribution of the noise, multiplied by 2^n (since we divide by 2^n when computing the expectation):

$$N_\tau^B = \mathbb{1}_B * f,$$

where $f(x) = 2^n \Pr[N_\tau(\mathbf{1}) = x]$. Here $\mathbf{1}\in\mathbb{B}^n$ is the unit of the group, i.e., $\mathbf{1} = (1,1,\dots,1)$. The Fourier coefficient $\widehat{f}_S$ is easy to compute:

$$\widehat{f}_S = \langle f,\chi_S\rangle = \mathop{\mathbb{E}}_{x\in\mathbb{B}^n} f(x)\chi_S(x),$$

and both functions f and χ_S are products of functions depending on one coordinate:

$$f(x_1,\dots,x_n) = g(x_1)\cdots g(x_n)$$

where $g(1) = 2-2\tau$ and $g(-1) = 2\tau$, and

$$\chi_S(x_1,\dots,x_n) = \chi_1(x_1)\cdots\chi_n(x_n),$$

where χ_i is constant 1 if $i\notin S$, and $\chi_i(x) = x$ for $i\in S$. Due to independence, the expectation of the product is a product of expectations; they are 1 for $i\notin S$ and $1-2\tau$ for $i\in S$, so

$$\widehat{f}_S = (1-2\tau)^{\#S}$$

In other terms, noise (convolution with f) decreases the S-th coefficient of the Fourier transform by multiplying it by $(1-2\tau)^{\#S}$. We need to apply noise to the indicator function of B that we denote by $b=\mathbb{1}_B$, and get a bound for the number of points where $b*f$ exceeds $1/n$.

Why $b * f$ cannot be relatively large (greater than $1/n$) on a large set A? We know that

$$b * f(x) = \sum_S (1 - 2\tau)^{\#S} \widehat{b}_S \chi_S(x).$$

This sum can be split into two parts: for "small" S, where $\#S < d$, and for "large" S, where $\#S \geqslant d$. Here d is some threshold to be chosen later in such a way that the first part (for small S) does not exceed, say $1/2n$ for all x. Then the second part should exceed $1/2n$ everywhere on A, and this makes the L_2-norm of the second part large, while all coefficients in the second part are multiplied by a small factor $(1 - 2\tau)^d$.

How should we choose the threshold d? The coefficient $\widehat{b}_\varnothing$ equals $\mu(B)$, the uniform measure of B, and for all other coefficients we have $|\widehat{b}_S| \leqslant \mu(B)$. The size (the number of terms) in the first part is the number of sets of cardinality less than d, and is bounded by $\operatorname{poly}(n) 2^{nH(d/n)}$. Therefore, if we choose d in such a way that

$$\mu(B) \operatorname{poly}(n) 2^{nH(d/n)} \leqslant \frac{1}{2n},$$

we achieve our goal (the first part of the sum never exceeds $1/2n$).

Now the second part: compared to the same part of the sum for b, we have all coefficients multiplied by $(1 - 2\tau)^d$, so the L_2-norm of this part is bounded:

$$\|\text{second part}\|_2 \leqslant (1 - 2\tau)^d \|b\|_2 = (1 - 2\tau)^d \sqrt{\mu(B)}.$$

On the other hand, if the second part exceeds $1/2n$ inside A, we have the lower bound:

$$\|\text{second part}\|_2 \geqslant \sqrt{\mu(A)}/2n$$

In this way we get

$$\sqrt{\mu(A)}/2n \leqslant (1 - 2\tau)^d \sqrt{\mu(B)},$$

or

$$\mu(A) \leqslant 4n^2 (1 - 2\tau)^{2d} \mu(B)$$

where d is chosen in such a way that

$$\mu(B) \leqslant 2^{-nH(d/n)} / \operatorname{poly}(n)$$

For $\#B = 2^{\beta n}$ we have $H(d/n) \approx 1 - \beta$ and

$$\#A \leqslant (1 - 2\tau)^{2d} 2^{\beta n}$$

We see that the first term gives an exponentially small factor since d is proportional to n:

$$d/n \approx H^{-1}(1 - \beta)$$

(here $H^{-1}(\gamma)$ is the preimage of γ between 0 and 1/2). So we get the required bound for some $\alpha < \beta$ as promised.

3.2 Using hypercontractivity

We can get a better bound using two-function hypercontractivity inequality for uniform bits, whose proof can be found in [O'Donnell, 2014, chapter 10]:

Proposition 4 (Two-function hypercontractivity inequality). *Let $f, g : \mathbb{B}^n \to \mathbb{R}$, let $r, s \geqslant 0$, and assume $0 \leqslant 1 - 2\tau \leqslant \sqrt{rs} \leqslant 1$. Then*

$$\mathop{\mathbb{E}}_{\substack{x \in \mathbb{B}^n \\ y = N_\tau(x)}} [f(x)g(y)] \leqslant \|f\|_{1+r} \, \|g\|_{1+s}$$

Here the distribution of x is the uniform distribution in $\mathbb{B}^n$, and y is obtained from x by applying (independent) τ-noise: $y = N_\tau(x)$. The same distribution can be obtained in a symmetric way, starting from y. The notation $\|\cdot\|$ denotes L_p-norm:

$$\|u\|_p = (\mathbb{E}|u^p|)^{1/p}.$$

How do we apply this inequality? For an arbitrary set B we consider the set

$$A = \{x : \Pr[N_\tau(x) \in B] \geqslant \varepsilon\}.$$

Let a, b be the indicator functions of A and B. Then Proposition 4 gives

$$\mathbb{E}[a(x)b(y)] = \Pr[x \in A, y \in B] \geqslant \Pr[x \in A] \Pr[y \in B | x \in A] \geqslant \mu(A)\varepsilon.$$

Now we write down the hypercontractivity inequality (note that $\|\mathbb{1}_X\|_q = \mu(X)^{1/q}$):

$$\begin{aligned} \varepsilon\mu(A) &\leqslant \mu(A)^{1/(1+r)}\mu(B)^{1/(1+s)} \\ \log\varepsilon + \log\mu(A) &\leqslant \frac{\log\mu(A)}{1+r} + \frac{\log\mu(B)}{1+s} \\ \log\mu(A) &\leqslant \frac{1+r}{r(1+s)}\log\mu(B) - \frac{1+r}{r}\log\varepsilon. \end{aligned}$$

This is true for every r, s with $\sqrt{rs} \geqslant 1 - 2\tau$. To get the strongest bound we minimize the right hand side, so we use (for given r) the minimal possible value of $s = (1-2\tau)^2/r$:

$$\log\mu(A) \leqslant \frac{1+r}{r + (1-2\tau)^2}\log\mu(B) - \frac{1+r}{r}\log\varepsilon.$$

If $\varepsilon = 1/\operatorname{poly}(n)$, we can set $r \to 0$ at the appropriate rate, and we finally get:

$$\begin{aligned} \log\mu(A) &\leqslant \frac{1}{(1-2\tau)^2}\log\mu(B) \\ \log\#A &\leqslant -\Big((1-2\tau)^{-2} - 1\Big)n + (1-2\tau)^{-2}\log\#B + o(n) \end{aligned}$$

3.3 Exact bound: combinatorial statement

We need a statement of the following type: if $B \subset \mathbb{B}^n$ is of size at most $2^{\beta n}$, and ε is some threshold (not too small, say, $1/n$ or even $1/\operatorname{poly}(n)$), then the set A of all $x \in \mathbb{B}^n$ that get into B with probability at least ε after applying a τ-random noise, is of size at most $2^{\alpha n}$. We have shown two arguments of this type, but they have not optimal (too strong) assumptions about α, β and τ. Figure 1 shows the corresponding regions of parameters α, β.

In this section we prove the result in the optimal version (matching the example with balls, as we discussed in Section 1). The proof goes in two steps: first we assume that ε is very close to 1, and then "amplify" this result and extend it to small values of ε.

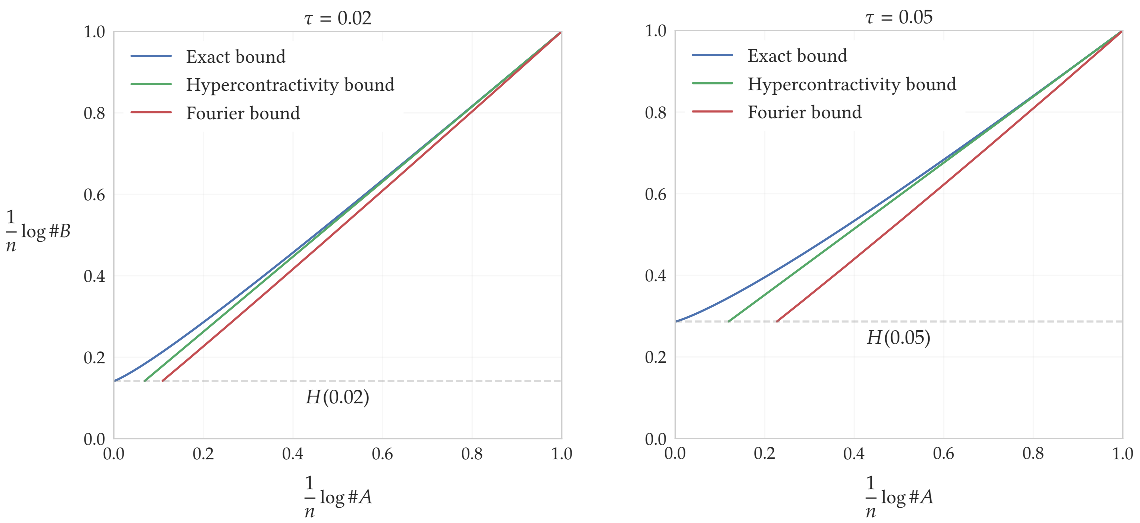

Figure 1: Bounds obtained with discussed techniques. The set of points not discarded by each bound consists of the points above the corresponding curve.

3.3.1 Proof for high probability range

First we show that *adding noise to a random variable increases its entropy.*

Let P be a random variable with values in $\mathbb{B}^n$, and P' be a noise version of P (we independently change every bit in the value of P with some probability τ). The first step[3] is to show that P' has larger entropy than P by providing some bound relating them (the bound depends on τ). It turns out that one can provide such a bound for $n = 1$ and this extends to all n ("one-letter characterization", "tensorization").

First, we need to consider a more general setting. Let X and Y be finite sets. Consider some stochastic transformation $T\colon X \to Y$: for every $x \in X$ we have some distribution $T(x)$ on Y. Then, for every random variable P with values in X, we may consider random variable $T(P)$ with values in Y. (In other words, we consider a random variable with values in $X \times Y$ whose marginal distribution on X is P and conditional distribution $Y|X$ is T.) For a fixed T (our main example is adding noise) we are interested in the relation between the entropies of P and $T(P)$ for arbitrary P. In other words, we consider the set of all pairs $(H(P), H(T(P)))$ for all possible X-valued random variables P. It is a subset of the rectangle $[0, \log\#X] \times [0, \log\#Y]$. We denote this set by $S(T)$. The following lemma shows that for a product of two independent transformations $T_1\colon X_1 \times Y_1$ and $T_2\colon X_2 \to Y_2$ this set can be bounded in terms of the correspoding sets for T_1 and T_2.

Lemma 1. *Let $T_1\colon X_1 \to Y_1$ and $T_2\colon X_2 \to Y_2$ be two stochastic transformations, and let $T_1 \times T_2\colon X_1 \times X_2 \to Y_1 \times Y_2$ be their product (independent transformations of both coordinates). Then every point (u, u') in $S(T_1 \times T_2)$ is above a sum of some point in $S(T_1)$ and some convex combination of points in $S(T_2)$.*

Here "above" means "can be obtained by increasing a second coordinate", and convex combination is linear combination with non-negative coefficients that have sum 1.

[3] In fact, the entropy increase for random variables is a corollary of the result about the complexity increase, for the same set of parameters. This can be proven in a usual way. We consider N independent copies of random variable P and independently apply noise to all of them. Then we write the inequality for the typical values of the complexities; in most cases they are close to the corresponding entropies, with $o(N)$ precision. Therefore, we get the inequality for entropies with $o(N)$ precision (for N copies) and $o(1)$ precision for one copy (the entropies are divided by N). As $N \to \infty$, we get rid of the additional term $o(1)$ and get an exact inequality for entropies.

Proof. Consider some random variable (P_1, P_2) with values in $X_1 \times X_2$; the components P_1 and P_2 can be dependent. Then

$$H(P_1, P_2) = H(P_1) + H(P_2|P_1).$$

This is the first coordinate of a pair in question; the second coordinate is the entropy of the variable $(T_1 \times T_2)(P_1, P_2)$; its components Q_1 and Q_2 are dependent and have (marginal) distributions $T_1(P_1)$ and $T_2(P_2)$. The second coordinate of the pair is then

$$H(Q_1, Q_2) = H(Q_1) + H(Q_2|Q_1).$$

We may consider all four variables P_1, P_2, Q_1, Q_2 as defined on the same space that is a product of three spaces: the space where (P_1, P_2) is defined, the space used in the stochastic transformation of P_1 and the space used in the stochastic transformation of P_2. Now we see that the pair we are interested it is a sum of two pairs:

$$(H(P_1, P_2), H(Q_1, Q_2)) = (H(P_1), H(Q_1)) + (H(P_2|P_1), H(Q_2|Q_1)).$$

The first pair $(H(P_1), H(Q_1))$ is in $S(T_1)$ by definition. The second pair, as we will show, is above $(H(P_2|P_1), H(Q_2|P_1))$. By definition, the conditional entropy with condition P_1 is a convex combination of conditional entropies with conditions $P_1 = x$ for all $x \in X$, and all pairs $(H(P_2|P_1 = x), H(Q_2|P_1 = x))$ are in $S(T_2)$, since for every x the distribution $(Q_2|P_1 = x)$ is obtained by applying T_2 to the distribution $(P_2|P_1 = x)$.

It remains to show that

$$H(Q_2|Q_1) \geqslant H(Q_2|P_1),$$

and this is because Q_1 and Q_2 are independent given P_1 (the difference $H(Q_2|Q_1) - H(Q_2|P_1)$ is equal to $I(Q_2 : P_1|Q_1) - I(Q_1 : Q_2|P_1)$, and the second term is zero due to the conditional independence. □

This lemma obviously generalizes for the product of several stochastic transformations. For the noise case in $\mathbb{B}^n$ we consider a product of n copies of "one-letter" transformation N_τ that maps 0 to 1 with probability τ and vice versa.

Lemma 2. *The set $S(N_\tau)$ is a curve in the unit square that starts at $(0, H(\tau))$ and ends at $(1, 1)$. This curve is increasing and convex.*

Proof of Lemma 2. This is an exercise in elementary calculus; still we provide the sketch of a proof. The curve in question is the image of the mapping

$$p \mapsto (H(p), H(p')),$$

where $p' = p + \tau - 2p\tau$, the probability to get 1 if we choose 1 with probability p and then change the result with probability τ (independently). The point p' divides the interval $[p, 1/2]$ as $2\tau : (1-2\tau)$. When p increases with constant speed from 0 to 1/2, the point p' also increases with constant speed from τ to 1/2, and the point $(H(p), H(p'))$ moves from left to right starting at $(0, \tau)$ and finishing at $(1, 1)$ (when $p = 1/2$, we have $p' = 1/2$). Then the curve is reversed, so we consider only $p \in (0, 1/2)$. To show that the curve is convex, we need to check that it slope increases from left to right (as p increases). Both points p and p' move with constant speeds, so the slope is proportional to the ratio $H'(p')/H'(p)$, where $H(p) = -p \log p - (1-p) \log(1-p)$.

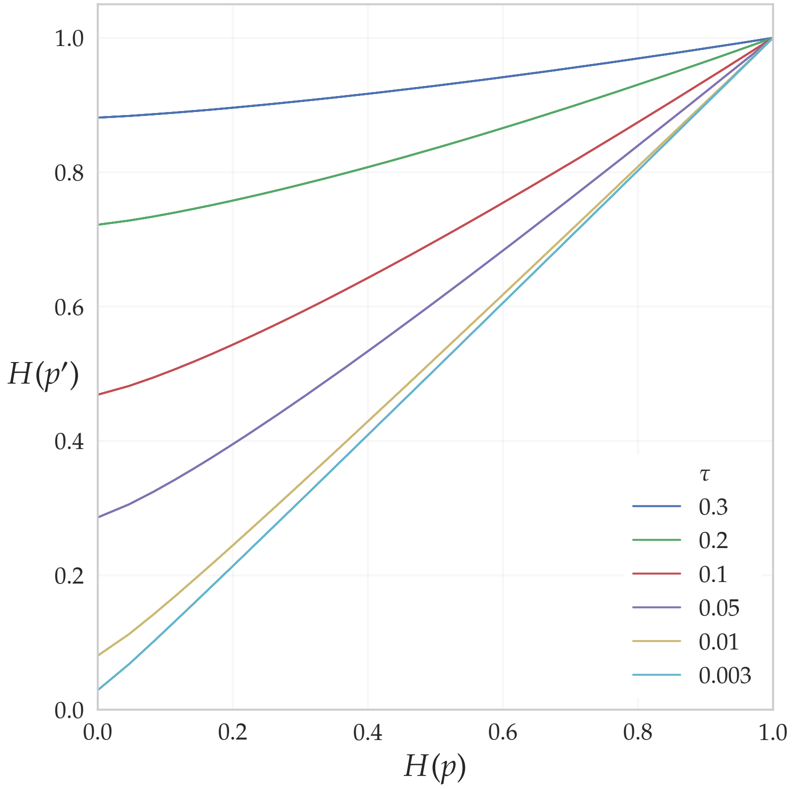

Figure 2: The boundary curve for N_τ for different values of τ.

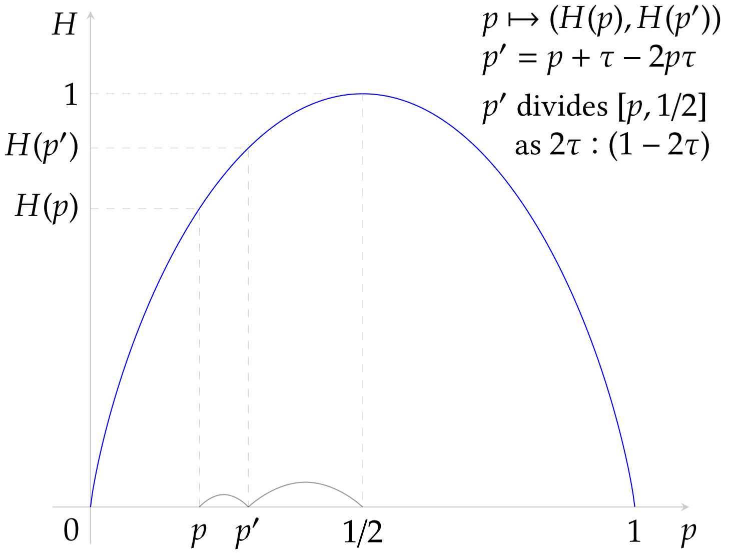

Figure 3: Obtaining a point $(H(p), H(p'))$ on the curve.

To compute the derivative $H'(p)$, we may replace the binary logarithms by natural ones (this does not change the ratio of derivatives). The derivative of $p\ln p$ is $\ln p + 1$, so

$$H'(p) = -\ln p - 1 + \ln(1-p) + 1 = \ln\left(\frac{1-p}{p}\right).$$

For computations, it is convenient to shift the origin and let $p = \frac{1}{2} + u$, then

$$H'(u) = \ln\left(\frac{1-2u}{1+2u}\right).$$

In this new coordinates p' corresponds to u' that is proportional to u, i.e., $u' = cu$, where c is a constant ($c = 1 - 2\tau$). We need to show that $H'(u')/H'(u)$ increases as u increases from $-1/2$ to 0. Letting $u = -v/2$, we need to show that

$$\ln\left(\frac{1+cv}{1-cv}\right)/\ln\left(\frac{1+v}{1-v}\right)$$

increases as v decreases from 1 to 0. Using the series

$$\ln\left(\frac{1+v}{1-v}\right) = \ln(1+v) - \ln(1-v) = 2(v + \frac{1}{3}v^3 + \frac{1}{5}v^5 + \ldots),$$

we can rewrite our statement as follows: the ratio

$$\frac{c\cdot v + c^3\cdot\frac{1}{3}v^3 + c^5\cdot\frac{1}{5}v^5 + \ldots}{v + \frac{1}{3}v^3 + \frac{1}{5}v^5 + \ldots}$$

decreases as v increases from 0 to 1. This ratio is a center of gravity for points having (decreasing) coordinates $c, c^3, c^5, \ldots$ and masses $v, \frac{1}{3}v^3, \frac{1}{5}v^5, \ldots$. When v increases, the proportion of masses is shifted to the right, and the sequence of coordinates decreases, so the center of gravity moves to the left as required. To say it a bit more formally, we note that the ratio of the first mass (v) and the rest ($\frac{1}{3}v^3 + \frac{1}{5}v^5 + \ldots$) decreases as v increases, so the center of gravity become closer to the center of gravity for the system without the first mass, and the latter also moves to the right for similar reason. To make the induction formal, we need to prove the similar statement for finitely many masses and then consider the limit.[4] □

Lemma 2 shows that the set of points of the unit square above $S(N_\tau)$ is convex. Therefore, applying Lemma 1 for the noise case, we do not need convex combinations: one point in each set $S(T_i)$ is enough. Note also that for N copies we have a sum of N points above $S(N_\tau)$, and dividing this sum by N, we get a point in $S(N_\tau)$. We get the following relationship between the entropy of an arbitrary random variable P in $\mathbb{B}^n$ and its noisy version.

Proposition 5. *Let P be arbitrary random variables with values in $\mathbb{B}^n$, and let P' be its noisy version obtained by applying N_τ independently to each bit in P. Choose p in such a way that $H(P) = nH(p)$. Then consider $q = N(p,\tau)$, the probability to get 1 if we apply N_τ to a variable that equals 1 with probability p. Then $H(P') \geqslant nH(q)$.*

[4] In general, we make use the following monotonicity statement: *if the coordinates of points are $x_1 > x_2 > \ldots > x_n$ and the masses $m_1, \ldots, m_n$ are changed in such a way that new masses m'_i satisfy the inequality $m'_j/m'_i > m_j/m_i$ for $j > i$, then center of gravity moves to the left after the change.* This can be easily proven by induction over n, following the scheme explained above.

In other words, for a fixed entropy $H(P)$ the minimal entropy of P' is achieved for Bernulli distribution $P = B_p$ for a suitably chosen p. Note that this is a Shannon information theory counterpart of our main result about increasing complexity by random noise. However, for the Kolmogorov complexity version it is not enough, and we need additional arguments amplifying the error probability. This is the topic of the next section.

3.3.2 Amplification: blowing-up lemma

Now we can apply the inequality for entropies to get some bound for the set size. Recall that we consider sets $A, B \subset \mathbb{B}^n$, and we know that for each $x \in A$ the probability of the event "$N_\tau(x) \in B$" is at least ε. We want to get the upper bound for the size of A in terms of the size of B and other parameters (ε and τ). We start by considering the regime when ε is close to 1.

Consider a random variable P that is uniformly distributed in A. Its entropy is $\log \#A$. We know that the entropy increases when we apply the noise to P and consider $N_\tau(P)$. On the other hand, with probability ε (close to 1, as we now assume) we have $N_\tau(P)$ in B. Indeed, this is the case for every $x \in A$, so it is also true for a randomly chosen x. So we can encode $N_\tau(P)$ in the following way:

- one bit (flag) says whether $N_\tau(P)$ is in B;
- if yes, then $\log \#B$ bits are used to encode an element of B;
- otherwise n bits are used to encode the value of $N_\tau(P)$ (trivial encoding).

Computing the average length of this code, we get the bound

$$H(N_\tau(P)) \leqslant 1 + \varepsilon \log \#B + (1 - \varepsilon)n. \qquad (*)$$

This bound makes sense if ε is close to 1. It does not give us much *per se* (recall that we are interested in the opposite regime, when $\varepsilon = 1/n$), but then we use some kind of amplification. The inequality (*) gives the desired bound (in the same region as for the random variables) if $\varepsilon = 1 - o(1)$ as $n \to \infty$. Therefore it is enough to extend our argument that was valid for the case $\varepsilon = 1 - o(1)$, to a more difficult case $\varepsilon = 1/\operatorname{poly}(n)$ (in fact, we need $\varepsilon = 1/n$, but the argument is the same for all polynomials).

The idea of this improvement is very simple. Consider for some d (depending on n, see below) the Hamming d-neighborhood of B. Let us denote it by B_d. We will show two things:

- B_d is not much bigger than B (for suitable d);
- if the probability to get into B after applying noise is at least $1/\operatorname{poly}(n)$ for every $x \in A$, then the probability to get into bigger set B_d is close to 1 for every $x \in A$.

The first statement uses the trivial bound: $\#B_d$ is bounded by $\#B$ multiplied by the size of the Hamming ball in $\mathbb{B}^n$ of radius d. We need that this additional factor does not change α and β asymptotically, so the value of d (for a given n) should be chosen in such a way that the size of the Hamming ball is $2^{o(n)}$. Since the size of the Hamming ball of radius d is $2^{nH(d/n)}$ and the entropy function $H(p)$ converges to 0 as $p \to 0$, it is enough to have $d = o(n)$.

What do we need for the second statement? We have some x and know that $\Pr[N_\tau(x) \in B] > 1/\operatorname{poly}(n)$. We need to show that $\Pr[N_\tau(x) \in B_d]$ converges to 1, for suitable chosen values of d. In fact, x does not matter here, we may assume that $x = 0...0$ (flipping bits in x and B simultaneously). For that we use the following property of Bernoulli distribution with parameter τ: if some set B has probability not too small (at least $1/\operatorname{poly}(n)$) according to this distribution, then its neighborhood B_d has probability close to 1. This statement is needed for $d = o(n)$.

Such a statement is called *blowing-up lemma* in [Ahlswede et al., 1976]. There are several (and quite different) ways to prove statements of this type. The original proof in [Ahlswede et al., 1976] used a result of Margulis from [Margulis, 1974] that says that the (Bernoulli) measure of a boundary of a set $U \subset \mathbb{B}^n$ is not too small compared to the measure of a boundary of a ball of the same size. Iterating this statement (a neighborhood is obtained by adding boundary layer several times), we get the lower bound for the measure of the neighborhood. Then Marton [Marton, 1986] suggested another proof based on the information-theoretical considerations that introduced transportation cost inequalities for bounding measure concentration. This proof provides $d = O(\sqrt{n \log n})$.

Let us state the blowing-up lemma in a slightly more general version than we need. Let $X_1, \ldots, X_n$ be (finite) probability spaces. Consider the space $X = X_1 \times \ldots \times X_n$ with the product measure μ (so the coordinates are independent).

Proposition 6 (Blowing-up lemma). *Let B be some subset of X of non-zero measure $\mu(B)$. Denote by B_r the r-neighborhood of B, i.e. the set of points that can be obtained from points in B by changing at most r coordinates. Then, if $r \geqslant \sqrt{(n/2)\ln(1/\mu(B))} + t$ for $t \geqslant 0$, the following bound holds:*

$$\mu(B_r) > 1 - \exp\left(-\frac{2t^2}{n}\right)$$

Remark 6. Let us check that the blowing-up lemma is enough for our purposes. We apply it to Boolean cube (so $X_i = \mathbb{B}$ for all i), and the noise distribution (which is a product distribution). If the probability of B is at least $1/n$ (or $1/\operatorname{poly}(n)$), then the expression $\sqrt{(n/2)\ln(1/\mu(B))}$ is $O(\sqrt{n \log n})$. If we let t be also of the same order, say, $t = \sqrt{cn \log n}$ for some constant c, we get $\mu(B_r) > 1 - \exp(-2c \log n) = 1 - n^{-2c}$, so $\mu(B_r)$ converges to 1 (rather fast, but this is not important). At the same time $r = O(\sqrt{n \log n}) = o(n)$ for this choice of t, as we needed.

So it remains to prove the blowing-up lemma to finish the proof of Theorem 2.

Remark 7. In fact, the blowing-up lemma is more symmetric than it looks at first. We claim that the complement of B_r is small if r is large enough. We can state it differently: if two sets B, B' are not too small, then the distance between B and B' is small. The exact statement is

$$d(B, B') \leqslant \sqrt{(n/2)\ln(1/\mu(B))} + \sqrt{(n/2)\ln(1/\mu(B'))}.$$

To get the original statement, assume that B' is the complement of B_r. Then the distance exceeds r, therefore

$$\sqrt{(n/2)\ln(1/\mu(B))} + \sqrt{(n/2)\ln(1/\mu(B'))} > r,$$

and

$$\sqrt{(n/2)\ln(1/\mu(B'))} > r - \sqrt{(n/2)\ln(1/\mu(B))} \geqslant t.$$

Therefore, $\mu(B') < \exp(2t^2/n)$, and we get the desired inequality (recall that B' is the complement of B_r).

To prove the blowing-up lemma, we use McDiarmid's inequality:

Proposition 7 (McDiarmid's inequality, [McDiarmid, 1989]). *Consider a function $f \colon X_1 \times \ldots \times X_n \to \mathbb{R}$. Assume that changing the i-th coordinate changes the value of f at most by some c_i:*

$$|f(x) - f(x')| \leqslant c_i,$$

if x and x' coincide everywhere except for the i-th coordinate.Then

$$\Pr[f - \mathbb{E} f \geqslant z] \leqslant \exp\left(-\frac{2z^2}{\sum_{i=1}^{n} c_i^2}\right)$$

for arbitrary $z \geqslant 0$.

Here the probability and expectation are considered with respect to the product distribution μ (the same as in the blowing-up lemma, see above). This inequality shows that f cannot be much larger that its average on a big set. Applying this inequality to $-f$, we get the same bound for the points where the function is less than its average by c or more.

We postpone the proof of McDiarmid's inequality to the end of the section. Now let us show why it implies the blowing-up lemma (in the symmetric version).

Proof of the blowing-up lemma. Let $f(x) = d(x, B)$ be the distance between x and B, i.e., the minimal number of coordinates that one has to change in x to get into B. This function satisfies the bounded differences property with $c_i = 1$, so we can apply McDiarmid's inequality to it. Let c be the expectation of f. Then on B the function is below that expectation at least by c, so

$$\mu(B) \leqslant \exp\left(-\frac{2c^2}{n}\right), \quad \text{or} \quad c \leqslant \sqrt{(n/2)\ln(1/\mu(B))}$$

On B' the function f is at least $d(B, B')$, so it exceeds its expectation at least by $d(B, B') - c$, therefore McDiarmid's inequality gives

$$d(B, B') - c \leqslant \sqrt{(n/2)\ln(1/\mu(B'))},$$

and it remains to add the last two inequalities. □

4 Acknowledgments

The work on this paper was done in collaboration with Alexander Shen, Peter Gács and Ilya Razenshteyn.

References

[Ahlswede et al., 1976] Ahlswede, R., Gács, P., and Körner, J. (1976). Bounds on conditional probabilities with applications in multi-user communication. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34(2):157–177.

[Buhrman et al., 2005] Buhrman, H., Fortnow, L., Newman, I., and Vereshchagin, N. K. (2005). Increasing kolmogorov complexity. In Diekert, V. and Durand, B., editors, *STACS 2005, 22nd Annual Symposium on Theoretical Aspects of Computer Science, Stuttgart, Germany, February 24-26, 2005, Proceedings*, volume 3404 of *Lecture Notes in Computer Science*, pages 412–421. Springer.

[Frankl and Füredi, 1981] Frankl, P. and Füredi, Z. (1981). A short proof for a theorem of harper about hamming-spheres. *Discrete Mathematics*, 34(3):311–313.

[Hoeffding, 1963] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30.

[Li and Vitányi, 2008] Li, M. and Vitányi, P. M. B. (2008). *An Introduction to Kolmogorov Complexity and Its Applications, Third Edition.* Texts in Computer Science. Springer.

[Margulis, 1974] Margulis, G. A. (1974). Veroyatnostniye characteristiki grafov s bolshoy svyaznostyu. *Problems of Information Transmission*, 10(2):174–179.

[Marton, 1986] Marton, K. (1986). A simple proof of the blowing-up lemma. *IEEE Trans. Information Theory*, 32(3):445–446.

[McDiarmid, 1989] McDiarmid, C. (1989). *On the method of bounded differences*, page 148–188. London Mathematical Society Lecture Note Series. Cambridge University Press.

[O'Donnell, 2014] O'Donnell, R. (2014). *Analysis of Boolean Functions.* Cambridge University Press.

[Shen et al., 2017] Shen, A., Uspensky, V. A., and Vereshchagin, N. (2017). *Kolmogorov complexity and algorithmic randomness*, volume 220. American Mathematical Soc.