On The Algorithmic Law of Large Numbers

And the Rarity of Sequences that Don't Conform to the LLN

Duarte Maia

January 1, 2025

1 Introduction

This document is a distillation of my reading of Proposition 3.2.13 in [1]: Each ML-random set satisfies the law of large numbers. In the entirety of this document, "random" will always be used to mean "Martin-Löf random".

This document is intended as a proof of the law of large numbers for random sets, which we state in its fullness for posteriority:

Theorem 1. If A is a random set, then A satisfies the law of large numbers, in the sense that $\frac{\#A\restriction n}{n} \to \frac{1}{2}$.

The proof we outline below has two parts. First, we prove that if a property Q is "likely", in the sense that the probability that a randomly-chosen sequence of size n has property Q gets closer to 1 as $n \to \infty$ at a rate to be determined below, then any random sequence will eventually satisfy the property Q. Then, we apply probability-theoretic tools to show that the property of "satisfying the law of large numbers up to error ε" is likely in this sense, implying by the definition of limit that Theorem 1 holds. That said, this document is intended to be mostly self-contained, at least when it comes to probability. The reader is nevertheless assumed to be comfortable with the notion of Martin-Löf randomness and prefix-free complexity.

2 Random Implies Not Rare

Let P be a property of strings. We will show that if P is a (computable enough) "rare" property, in a sense to be defined shortly, then for any random sequence A almost all prefixes of A satisfy $\neg P$. More precisely, for all but finitely many $n \in \mathbb{N}$ we have $\neg P(A\restriction n)$.

The strategy that we will follow will be the following. We describe the prefixes of A, or at least the ones that satisfy P, by encoding their length n, followed by instructions to print out the sequences of this length that satisfy P, followed by saying "the string I want is the k-th one". If P is rare enough, then k will be small enough that this encoding is better than linear.

Let's do some back-of-the envelope calculations to figure out how rare P should be. Let's say that we want to encode the string σ in this manner. Say that its length is n, and in some previously-fixed enumeration of the strings of length n satisfying P, σ comes in k-th place. Then, let us encode the pair $\langle n, k\rangle$ in an efficient prefix-free way.

- The most direct way would be to encode using unary, which yields $n + 1 + k + 1$, which will never be smaller than n.
- Let us instead encode in binary, using unary to encode the order of magnitude. This yields[1] $2\log n + 2\log k$, which grows slower than n so long as $2\log k \leq n - 2\log n - u_n$, with u_n some sequence converging

[1] Ignoring some constant terms arising from the need to use separators, a habit which we will follow henceforth.

to infinity. This will be satisfied so long as P is rare enough that the number N_n of strings of length n satisfying P satisfies (for example) $2\log N_n \leq n - (1+\varepsilon)\log n$, or equivalently

$$N_n^2 \leq 2^n / n^{1+\varepsilon}. \tag{1}$$

This is a really bad bound. To see this, let us rephrase it in terms of density, or probability. Let $p_n = N_n/2^n$ be the proportion of strings of size n that satisfy p. Then, Equation (1) may be rewritten as

$$p_n \leq 2^{-n/2} / n^{\frac{1}{2}+\varepsilon'}.$$

In other words, we need that the proportion of strings satisfying P decreases exponentially. This is no good. Fortunately, we can notice that the exponential term comes from the term $2\log k$ in the size of the encoding. If we encode a little smarter, like, say, encoding the order of magnititude itself in binary, we will get a polynomial bound.

- Let us encode n in binary as before, and let us now encode k in binary, with the order of magnitude itself in binary. This yields $2\log n + \log k + 2\log\log k$. Thus, if N_n is as above, we now want the bound

  $$2\log n + \log N_n + 2\log\log N_n \leq n - u_n$$

  for some $u_n \to \infty$. This suggests, though it will not be sufficient, to assume the bound $\log N_n \leq n - 2\log n$. Once we do, we will get

  $$2\log n + \log N_n + 2\log\log N_n \leq n + 2\log(n - 2\log n),$$

  which is in fact not growing slower than n. However, if we subtract another two log terms from our bound, we get something much more promising:

  $$\log N_n \leq n - 4\log n \implies 2\log n + \log N_n + 2\log\log N_n \leq n - 2\log\left(\frac{n}{n-4\log n}\right).$$

  Unfortunately, $\frac{n}{n-4\log n} \to 1$ and so $\log\left(\frac{n}{n-4\log n}\right)$ converges to 0, not to ∞. Thus, this is barely not enough.

- Let us now try to take the encoding of k one step deeper, giving the bound $2\log n + \log k + \log\log k + 2\log\log\log k$. Now, if $N_n \leq n - 4\log n$ we get

  $$2\log n + \log N_n + \log\log N_n + 2\log\log\log N_n \leq n - \log\left(\frac{n^2}{(n-4\log n)(\log(n-4\log n))^2}\right), \tag{2}$$

  and we can now verify that the log term does in fact converge to infinity. This proves a preliminary version of our result:

Proposition 2. Let P be a c.e. property of strings that is rare in the sense that, for every n,

$$\frac{\text{number of strings of size } n \text{ that satisfy } P}{2^n} \leq \frac{1}{n^4}.$$

Then, if A is a random sequence, only finitely many prefixes of A satisfy P.

Proof: The paragraphs that precede this proposition provide a way to encode string satisfying P in a sublinear way. Thus, if infinitely many prefixes $A\upharpoonright n_i$ of A satisfied P, we would have $K(A\upharpoonright n_i) \leq n_i - u_{n_i}$, which implies by the K-complexity definition of 1-randomness that A is nonrandom. ■

This bound can be improved. If we also encode n in a smarter way, using $\log n + 2\log\log n$ bits, something similar to Equation (2) may be recovered assuming only $N_n \leq n - 3\log n$. There is also no harm in adding a constant term. In other words:

Proposition 3. Let P be a c.e. property of strings such that, for some constant c, for every $n \in \mathbb{N}$ we have $\frac{\text{number of strings of size } n \text{ that satisfy } P}{2^n} \leq \frac{c}{n^3}$. Then, if A is a random sequence, only finitely many prefixes of A satisfy P.

3 Most Strings Satisfy the Law of Large Numbers

Let A be a random set. We want to show that the Law of Large Numbers holds for A, that is,

$$\lim_{n\to\infty} \frac{\#A\restriction n}{n} = \frac{1}{2}.$$

By expanding the definition of limit, this boils down to saying that, for any $\varepsilon > 0$, for all but finitely many n we have $\left|\frac{\#A\restriction n}{n} - \frac{1}{2}\right| < \varepsilon$. In light of Proposition 3, this suggests proving that, for fixed $\varepsilon > 0$, the proportion of strings σ of size n that satisfy $\left|\frac{\#\sigma}{n} - \frac{1}{2}\right| \geq \varepsilon$ decreases at least cubically, where $\#\sigma$ denotes the number of ones in σ. So let's establish some notation to help us approach the problem.

Definition 4. For $p \in \mathbb{R}$ and $n \in \mathbb{N}$, define

$$A_{np} = \{\, \sigma \in 2^n \mid \#\sigma \leq pn \,\},$$

and define analogously B_{np} as the set of strings of size n with $\#\sigma \geq pn$.

The quantity we intend to bound is the amount of strings satisfying $\left|\frac{\#\sigma}{n} - \frac{1}{2}\right| \geq \varepsilon$, or equivalently

$$\#\sigma \leq \tfrac{1}{2} - \varepsilon \ \vee \ \#\sigma \geq \tfrac{1}{2} + \varepsilon.$$

Thus, we intend to bound the size of $A_{n\left(\frac{1}{2}-\varepsilon\right)} \cup B_{n\left(\frac{1}{2}+\varepsilon\right)}$. Moreover, the operation of swapping zeros on a string for ones and vice-versa provides a bijection between these two sets, and so it suffices to get a good upper bound on the quantity

$$P_{np} = \frac{\#A_{np}}{2^n},$$

for $p < \frac{1}{2}$. In this section, we will provide four ways to obtain upper bounds, of which one is insufficient, and another merely heuristic.

3.1 Failed Approach: Chebyshev

The first attempt at a bound on the size of A_{np} might be via the use of Chebyshev's inequality (though I believe probabilists refer to what I'm about to state as Markov's inequality). We state it for the particular case of finite measure spaces, which is the only case for which we will use it.

Theorem 5 (Chebyshev). If $f\colon X \to [0, \infty[$, for X a finite set, and $a \geq 0$ is a real number, we have

$$\sum_{x\in X} f(x) \geq a \times \#\{x \mid f(x) \geq a\}.$$

A first try would be to apply it to the function

$$\begin{aligned} f\colon 2^n &\to \mathbb{N} \\ \sigma &\mapsto \#\sigma \end{aligned}$$

getting the inequality: $\sum_{\sigma\in 2^n} f(\sigma) \geq \#\{\sigma \mid f(\sigma) \geq qn\} \times qn$. The left-hand side is seen by combinatorial arguments to equal $n2^{n-1}$, because the involution $\sigma \mapsto \sigma^*$ that swaps zeros and ones divides the space 2^n into orbits of size two, for which $f(\sigma) + f(\sigma^*) = n$. The right-hand side is $qn\,\#B_{nq}$. Applying this to $q = 1 - p$ and using the fact that $\#A_{np} = \#B_{n(1-p)}$, we get the bound

$$\#A_{np} \leq \frac{1}{2(1-p)} 2^n.$$

This is not a very good bound. We want the proportion of elements in $\#A_{np}$ to decrease cubically, and we failed to even show that it goes to zero.

A better bound – and this is the one that probabilists call Chebyshev's inequality – consists of applying Theorem 5 to the function

$$\begin{aligned} g\colon 2^n &\to [0,\infty[\\ \sigma &\mapsto (\#\sigma - n/2)^2, \end{aligned}$$

with $a = n^2\varepsilon^2$This yields the inequality

$$\#A_{n\left(\frac{1}{2}-\varepsilon\right)} + \#B_{n\left(\frac{1}{2}+\varepsilon\right)} \leq \frac{1}{\varepsilon^2} \sum_{\sigma\in 2^n} g(\sigma).$$

It remains to compute $\sum g(\sigma)$. For this task, it is probably best to change to purely probabilistic methods. Indeed, to calculate $\frac{\sum g(\sigma)}{2^n}$ is to compute the variance of $X_1 + \cdots + X_n$, with the X_i being independent random variables with value 0 or 1 with probability $\frac{1}{2}$. We have the formula from probability

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y),$$

for X and Y independent variables. Thus, $\mathrm{Var}(X_1 + \cdots + X_n) = n\mathrm{Var}(X)$, for X a zero-one coin flip, and $\mathrm{Var}(X) = \frac{1}{4}$. Thus,

$$\frac{\#A_{n\left(\frac{1}{2}-\varepsilon\right)} + \#B_{n\left(\frac{1}{2}+\varepsilon\right)}}{2^n} \leq \frac{1}{4\varepsilon^2 n}.$$

This is a better bound, but it still falls short, decreasing at a rate of $1/n$ and not the required $1/n^3$.

3.2 Heuristic Approach

We now present a heuristic approach that suggests that we can do much better than $1/n^3$. We will make this heuristic approach rigorous in Section 3.3.

A direct counting argument gives us a reasonably explicit expression for $\#A_{np}$:

$$\#A_{np} = \sum_{k\leq np} \binom{n}{k}.$$

We approximate $\binom{n}{k}$ by a smooth expression, and approximate the sum by the integral of this expression. To obtain an approximation for $\binom{n}{k}$, we also apply a similar technique.

First, we recall and sketch Stirling's approximation for the factorial. Note that $\log n! = \log n + \cdots + \log 1$, which we approximate by $\int_1^n \log x\,\mathrm{d}x = n\log n + n - 1$. Thus, $\log n! \approx n\log n + n - 1$, and so

$$n! \approx \frac{n^n \mathrm{e}^n}{\mathrm{e}}.$$

From this and the expression for the binomial, we obtain the approximation

$$\binom{n}{k} \approx \frac{1}{\mathrm{e}} \frac{n^n}{k^k (n-k)^{n-k}}.$$

Thus, we can approximate

$$\#A_{np} = \sum_{k\leq np} \binom{n}{k} \approx \int_0^{np} \frac{1}{\mathrm{e}} \frac{n^n}{k^k (n-k)^{n-k}}\,\mathrm{d}x,$$

which by a change of variables to $y = x/n$ is given by

$$\#A_{np} \approx \frac{n}{\mathrm{e}} \int_0^p \left(\frac{1}{y^y (1-y)^{1-y}} \right)^n \, \mathrm{d}y. \tag{3}$$

This integral is not amenable to direct computation, but it can be approximated. Let us look at plots of the integrand, for large values of n.

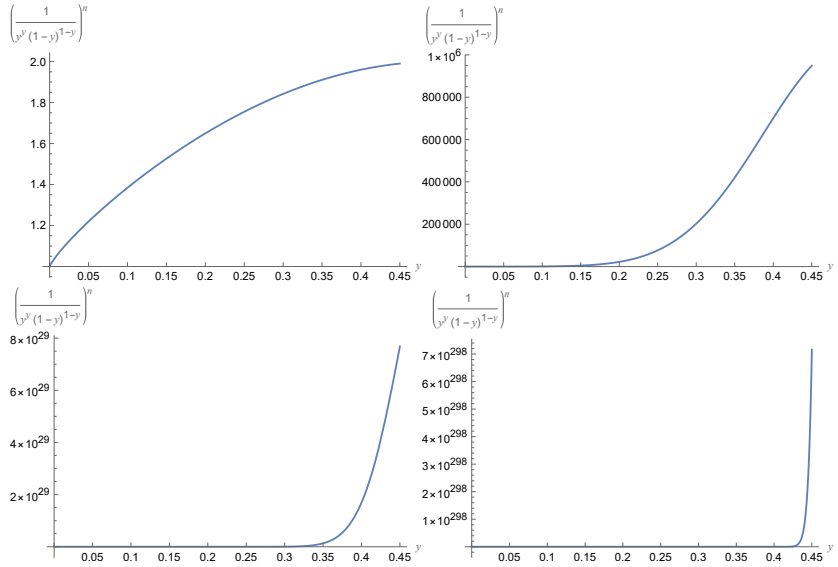

Figure 1: A plot of $\left(\frac{1}{y^y (1-y)^{1-y}} \right)^n$, for $n = 1$, 20, 100, and 1000, for $y \in [0, 0.45]$.

We can see from the figures that, as $n \to \infty$, only the rightmost part of the graph is significant. We can approximate the graph by a right triangle, and its integral by the area of this triangle.

Remark 6. The above approximation is only valid for $p < \frac{1}{2}$!

Set $h_n(x) = \left(\frac{1}{y^y (1-y)^{1-y}} \right)^n$. Then, for large values of n, the triangle we refer to in the previous paragraph has height given by $h_n(p) = \left(\frac{1}{p^p (1-p)^{1-p}} \right)^n$, and its slope is

$$m = h_n'(p) = n \left(\frac{1}{p^p (1-p)^{1-p}} \right)^{n-1} h_1'(p),$$

and a computation yields $h_1'(y) = \frac{1}{y^y (1-y)^{(1-y)}} \log\left(\frac{1-y}{y} \right)$. Putting it all together, and using the formula for the area of a right triangle: $A = \frac{y^2}{2m}$, we get

$$\#A_{np} \approx \frac{n}{\mathrm{e}} \frac{\left(\frac{1}{p^p (1-p)^{1-p}} \right)^{2n}}{2n \left(\frac{1}{p^p (1-p)^{1-p}} \right)^n \log\left(\frac{1-p}{p} \right)} = \frac{1}{2\mathrm{e} \log\left(\frac{1-p}{p} \right)} \left(\frac{1}{p^p (1-p)^{1-p}} \right)^n.$$

Thus, the proportion of strings in A_{np} is approximated by

$$\frac{\#A_{np}}{2^n} \approx C (2 p^p (1-p)^{1-p})^{-n},$$

which, so long as $2p^p(1-p)^{1-p} > 1$, decreases geometrically, which is faster than cubically. This may be verified analytically (for $p \neq \frac{1}{2}$), but for now, the following plot corroborates this guess:

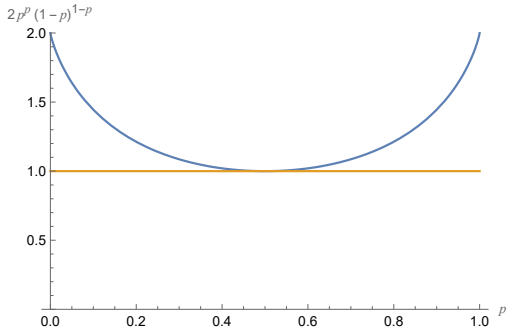

Figure 2: Plot of $2p^p(1-p)^{1-p}$, compared against $y = 1$.

This gives us our conclusion: *Not only does the proportion of strings in $\#A_{np}$ decrease to zero faster than cubically, but it does so exponentially, at a rate of $(2p^p(1-p)^{1-p})^{-n}$.*

In the following two sections, we verify that this indeed holds true.

3.3 First Successful Approach: Bounding Sums by Integrals

This section consists entirely of a rigorous approach to the contents of Section 3.2.

To recap: We seek to bound the value of

$$\#A_{np} = \sum_{k \leq np} \binom{n}{k}.$$

We begin by providing a bound for $\binom{n}{k}$. As before, we consider

$$\log \binom{n}{k} = \log n + \cdots + \log(n-k+1) - \log k - \cdots - \log 1. \tag{4}$$

Throughout, we will take recourse to the following principle:

Proposition 7. Let $f\colon [a, b+1] \to \mathbb{R}$ be an increasing function. Then,

$$\sum_{a \leq k \leq b} f(k) \leq \int_a^{b+1} f(x)\,\mathrm{d}x.$$

Proof: Use the lower Darboux sum given by the partition $(a, \lceil a \rceil, \lceil a \rceil + 1, \ldots, \lfloor b \rfloor, \lfloor b \rfloor + 1, b+1)$. ∎

An analogous idea, or applying Proposition 7 to $-f(-x)$ provides a lower bound.

Proposition 8. Let $f\colon [a-1, b] \to \mathbb{R}$ be an increasing function. Then,

$$\sum_{a \leq k \leq b} f(k) \geq \int_{a-1}^{b} f(x)\,\mathrm{d}x.$$

Applying Propositions 7 and 8 to Equation (4) (after disregading $\log 1 = 0$) yields the bound

$$\begin{aligned}
\log \binom{n}{k} &\leq \int_{n-k+1}^{n+1} \log(x)\,\mathrm{d}x - \int_1^k \log(x)\,\mathrm{d}x \\
&= (n+1)\log(n+1) + \cancel{(n+1)} - (n-k+1)\log(n-k+1) - \cancel{(n-k+1)} - k\log k - \cancel{k} + 1 \\
&= (n+1)\log(n+1) - (n-k+1)\log(n-k+1) - k\log k + 1,
\end{aligned}$$

providing the bound on the binomial:

$$\binom{n}{k} \leq \frac{\mathrm{e}\,(n+1)^{n+1}}{(n-k+1)^{n-k+1}k^k}.$$

Let us define the auxilliary function, which was seen to be useful in Section 3.2,

$$h(x) = \frac{1}{x^x(1-x)^{1-x}}.$$

We note that our bound on the binomial may be rewritten using it as recourse:

$$\binom{n}{k} \leq \mathrm{e}\,h\left(\tfrac{k}{n+1}\right)^{n+1}. \tag{5}$$

We may compute $h'(x) = h(x)\log\left(\frac{1-x}{x}\right)$, whose sign is the same as the sign of $\log\left(\frac{1-x}{x}\right)$. As a consequence $h(x)$, and hence the bound in (5), is increasing for $x < \frac{1}{2}$, resp. $k < \frac{n+1}{2}$. As such, we may apply Proposition 5 in conjunction with (5) to obtain

$$\sum_{k\leq np}\binom{n}{k} \leq \int_0^{np+1} \mathrm{e}\,h\left(\tfrac{x}{n+1}\right)^{n+1}\,\mathrm{d}x, \tag{6}$$

which by a change of variables to $y = \frac{x}{n+1}$ gives

$$\sum_{k\leq np}\binom{n}{k} \leq (n+1)\mathrm{e}\int_0^{\frac{np+1}{n+1}} h\,(y)^{n+1}\,\mathrm{d}x. \tag{7}$$

Now, let's recall that the quantity we are interested in is $\frac{1}{2^n}\sum_{k\leq np}\binom{n}{k}$. Unfortunately, since (7) is not as tidy as the heuristic bound (3), the computations are going to get a little ugly. Nevertheless, we will prove that $\frac{1}{2^n}\sum_{k\leq np}\binom{n}{k}$ decreases exponentially, which is faster than cubically and so will suffice to prove the Law of Large Numbers.

We divide the integral in 7 into two parts:

$$(n+1)\mathrm{e}\int_0^{\frac{np+1}{n+1}} h\,(y)^{n+1}\,\mathrm{d}x = \underbrace{(n+1)\mathrm{e}\int_0^{p} h\,(y)^{n+1}\,\mathrm{d}x}_{Q_1} + \underbrace{(n+1)\mathrm{e}\int_p^{\frac{np+1}{n+1}} h\,(y)^{n+1}\,\mathrm{d}x}_{Q_2}. \tag{8}$$

We start by bounding Q_2, as it is the easiest. We have the bound

$$Q_2 \leq (n+1)\mathrm{e}\left(\frac{np+1}{n+1} - p\right) h\left(\tfrac{np+1}{n+1}\right)^{n+1} = \mathrm{e}(1-p)h\left(\tfrac{np+1}{n+1}\right)^{n+1}.$$

Finally, since $\frac{np+1}{n+1} \to p$, $p < \frac{1}{2}$, and h is decreasing for $x < \frac{1}{2}$, we may choose n large enough that $\frac{np+1}{n+1}$ is always less than $p+\varepsilon$ for some small ε, whence $h\left(\frac{np+1}{n+1}\right) < h(p+\varepsilon) < 2$, and so $\frac{1}{2^n}Q_2$ decreases exponentially, completing the first part of the proof.

Let us now bound Q_1. In order to make the "approximate by a right triangle" argument from Section 3.2 rigorous, we employ the following lemma.

Lemma 9. Let $f\colon [0,1] \to [0,\infty[$ be a function whose left-derivative at 1, which we will refer to as $f'(1)$, exists and is nonzero. Then, the integral $\int_0^1 f(x)^n\,\mathrm{d}x$ is asymptotically equal to $\frac{f(1)}{nf'(1)}f(1)^n$. In other words,

$$\frac{\int_0^1 f(x)^n\,\mathrm{d}x}{\frac{f(1)}{nf'(1)}f(1)^n} \to 1.$$

Proof: By considering $\tilde{f} = \frac{1}{f(1)} f$, we may without loss of generality assume that $f(1) = 1$. This is not an essential part of the proof, but makes it notationally nicer, as now the question boils down to computing the limit of $n \int_0^1 f(x)^n \, \mathrm{d}x$, which we claim equals $f'(1)$.

We begin by considering the special case where

$$f(x) = \begin{cases} m(x-1)+1 & x \geq 1-\delta \\ c & x < 1-\delta \end{cases} \tag{9}$$

for some positive real number m, $0 \leq c < 1$, and small enough value of δ. In this case, the integral may be computed explicitly:

$$\int_0^1 f(x)^n \, \mathrm{d}x = (1-\delta)c^n + \int_{1-\delta}^1 (m(x-1)+1)^n \, \mathrm{d}x = (1-\delta)c^n + \left(\frac{1}{m(n+1)} - \frac{(1-\delta m)^{n+1}}{m(n+1)} \right).$$

Now, for δ small enough, we have $0 < 1 - \delta m < 1$, and so

$$n \int_0^1 f(x)^n \, \mathrm{d}x = n(1-\delta)c^n + \frac{n}{n+1} \left(\frac{1}{m} - \frac{(1-\delta m)^{n+1}}{m} \right) \to \frac{1}{m},$$

and so this special case is proven. Now, for the general case, we note that any function f that satisfies the assumptions of the problem statement, f may be bounded from above by a function of type (9) with $m = f'(1) + \varepsilon$, and from below by a function of type (9) with $m = f'(1) - \varepsilon$. Thus, any sublimit of $n \int_0^1 f(x)^n \, \mathrm{d}x$ is between $f'(1) - \varepsilon$ and $f'(1) + \varepsilon$, which by standard real analysis arguments implies that the limit exists and equals $f'(1)$. This completes the proof. ■

Remark 10. We didn't get a factor of $\frac{1}{2}$ in the assymptotic approximation, which means that in fact approximating by a right triangle is an underestimate by a factor of a half. It appears that we should have approximated by a rectangle instead, which is pretty weird!

Now equipped with Lemma 9, we can bound Q_1 easily. Indeed, by Lemma 9 we have

$$Q_1 = (n+1)\mathrm{e} \frac{h(p)}{nh'(p)} h(p)^n \times a_n,$$

with $a_n \to 1$. We see that $\frac{n+1}{n} \to 1$ also, and so the only thing that determines the growth of $Q_1$is the exponential term $h(p)^n$. To investigate this, note that we've calculated before that h is increasing (strictly) up to $\frac{1}{2}$, and moreover we can easily compute $h(1/2) = 2$. Thus, $h(p)$ is (for $p < 1/2$) *strictly* less than 2, and so we conclude that $\frac{1}{2^n} Q_1$ converges exponentially to zero, which is again faster than cubic, as desired.

Remark 11. While this proof is sufficient to get that $\frac{\# A_{np}}{2^n}$ goes to zero exponentially, it does not quite reach the exponential decay of rate $2p^p(1-p)^{1-p}$ that we expected from Section 3.2. We only obtained exponential decay with rate $2p^p(1-p)^{1-p} + \varepsilon$ for arbitrary ε. A better bound could be obtained by being more careful in step (6). Instead of bounding the whole sum by an integral, leave the last term out:

$$\sum_{k \leq np} \binom{n}{k} \leq \int_0^{np} \mathrm{e}\, h\left(\tfrac{x}{n+1}\right)^{n+1} \, \mathrm{d}x + \binom{n}{\lfloor np \rfloor}.$$

The first term is what we referred to in Equation (8) as Q_1, which we bounded by $c\,(p^p(1-p)^{1-p})^n$, and the second term we can bound using Equation (5) to get

$$\binom{n}{\lfloor np \rfloor} \leq \mathrm{e}\, h\left(\frac{\lfloor np \rfloor}{n+1} \right)^{n+1} \leq c'\, h(p)^n,$$

where we used the fact that h is increasing up to $\frac{1}{2}$ and $\frac{\lfloor np \rfloor}{n+1} \leq p$. This is a better bound than the bound we obtained for Q_2, and gives us an exponential bound with rate $h(p) = p^p(1-p)^{1-p}$ as desired.

Remark 12. I suspect that the same approach could be used to furnish a lower bound for $\frac{\# A_{np}}{2^n}$, and hence prove that the decay bound $(2p^p(1-p)^{1-p})^{-n}$ is tight up to multiplicative factors, but I have been unable to get it to work.

3.4 Second Successful Approach: Chernoff Bounds

This approach boils down to applying Theorem 5 to a well-chosen family of functions $\{f_t\}$, and keeping the best bound we obtain from this process. These functions are chosen as to exploit regularities of our specific scenario. Probabilistically speaking, what we have is a sum of independent random variables $X_1 + \cdots + X_n$, each of which is zero or one with probability $1/2$. A remarkable property of independent r.v.s is that, if X and Y are independent, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, which suggests taking an exponential to turn the sum into a product. Thus, we consider the bounds obtained by applying Chebyshev/Markov's inequality to $\lambda^{X_1+\cdots+X_n}$ for $\lambda > 0$, or equivalently to $\exp(t(X_1 + \cdots + X_n))$ for $t \in \mathbb{R}$. We will in fact only consider $t > 0$, because other choices for t do not preserve inequalities.

To begin the proof, let us recall that what we want to estimate is the proportion of strings in 2^n that lie in A_{np}. In the framework of the previous paragraph, this is the same as to estimate $\mathbb{P}[X_1 + \cdots + X_n \leq np]$, which is the same as $\mathbb{P}[\mathrm{e}^{tX_1} \ldots \mathrm{e}^{tX_n} \leq \mathrm{e}^{tnp}]$ for arbitrary $t > 0$. In turn, we apply Markov's inequality here to obtain

$$\mathbb{P}[\mathrm{e}^{tX_1} \ldots \mathrm{e}^{tX_n} \leq \mathrm{e}^{tnp}] \leq \frac{1}{\mathrm{e}^{tnp}} \mathbb{E}[\mathrm{e}^{tX_1} \ldots \mathrm{e}^{tX_n}] = \frac{\mathbb{E}[\mathrm{e}^{tX}]^n}{\mathrm{e}^{tnp}}, \tag{10}$$

where we used the fact that the X_i are independent, and hence the e^{tX_i} are independent. We also use X to denote an arbitrary r.v. with the same distribution as the X_i. Finally, we optimize the right-hand side of Equation (10). Using $\mathbb{E}[\mathrm{e}^{tX}] = \frac{1+\mathrm{e}^t}{2}$, what follows is a calculus exercise, whose conclusion is that the optimal value of t is $t = \log\left(\frac{p}{1-p}\right)$, which yields the bound

$$\mathbb{P}[X_1 + \cdots + X_n \leq np] = \mathbb{P}[\mathrm{e}^{tX_1} \ldots \mathrm{e}^{tX_n} \leq \mathrm{e}^{tnp}] \leq \left(\frac{1}{2} \frac{1 + \frac{p}{1-p}}{\left(\frac{p}{1-p}\right)^p}\right)^n = (2p^p(1-p)^{1-p})^{-n}. \tag{11}$$

This is already enough to prove the Law of Large Numbers for random sequences, as it provides us with exponential decay, which is faster than cubic decay.

To conclude this document, we briefly explain how to go from bound (11) to the Chernoff bounds outlined by Nies on page 101. The main step is to prove that $p^p(1-p)^{1-p} \geq \frac{1}{2}\mathrm{e}^{(p-\frac{1}{2})^2}$. This can be done by optimizing the logarithm of the quotient $\frac{p^p(1-p)^{1-p}}{\exp((p-\frac{1}{2})^2)}$, which is $p \log p + (1-p)\log(1-p) + (p-\frac{1}{2})^2$. Its derivative is $\log\left(\frac{p}{1-p}\right) - 2p - 1$, which can be seen to be negative for $p < \frac{1}{2}$ by using the bound $\log x \leq x - 1$, and positive for $p > \frac{1}{2}$ because it's odd about $1/2$. Thus, the global minimum of $\frac{p^p(1-p)^{1-p}}{\exp((p-\frac{1}{2})^2)}$ is at $p = \frac{1}{2}$, with value $1/2$, which proves the inequality $p^p(1-p)^{1-p} \geq \frac{1}{2}\mathrm{e}^{(p-\frac{1}{2})^2}$. Therefore, Equation 11 has as a consequence that

$$\mathbb{P}[X_1 + \cdots + X_n \leq np] \leq \mathrm{e}^{-n(p-\frac{1}{2})^2}.$$

This gives a bound on the size of A_{np}. The bound on the size of $B_{n(1-p)}$ is the same, and so we get Nies' bound

$$\mathbb{P}\left[\left|\frac{X_1 + \cdots + X_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right] \leq 2\mathrm{e}^{-n\varepsilon^2}.$$

References

[1] André Nies. *Computability and randomness*, volume 51 of *Oxford Logic Guides*. Oxford University Press, Oxford, 2009.