Pattern avoiding strings

Jacob Richey

November 2, 2023

Contents

1 Introduction

In this note we consider strings over an alphabet $\mathcal{A}$, typically $\mathcal{A} = \{0, 1\}$, (and possibly $\mathcal{A} = [q]$ for some positive integer q or $\mathcal{A} = \mathbb{N}$), conditioned on avoiding some pattern set S. This can mean a few different things. As a warm-up, we can take 'pattern' to mean 'substring,' i.e. take $S \subset \Omega$, where $\Omega_n = \mathcal{A}^n$ is the set of sequences of length n and $\Omega = \cup_n \Omega_n$, and write

$$\Omega_n(S) = \{\omega \in \Omega_n : \omega \text{ does not contain } s \text{ as a substring for any } s \in S\}. \tag{1.1}$$

Here 'substring' means 'consecutive subsequence:' 11 is a substring of 1101, but 111 is not. (Disallowing arbitrary subsequences to match S seems quite restrictive, but could be interesting too.) For example, with $\mathcal{A} = \{0, 1\}$ and $S = \{11, 1001\}$, we have

$$\Omega_4(S) = \{0000, 0001, 0010, 0100, 1000, 1010, 0101\} \tag{1.2}$$

Of course, these $\Omega_n(S)$ generate all possible events if any sets S are allowed: we have in mind 'small' sets S. We want to study random strings sampled from some measure on $\Omega_n(S)$, or if it makes sense, $\Omega(S)$, or $\Omega_\infty(S)$: two examples to keep in mind are an iid string of fixed length conditioned to have no substring in S, or a string generated with iid bits, one bit at a time, and stopped on containing some string in S as a substring. The natural limits for these objects are *shifts of finite type*. There is also work focusing on the expected hitting time of a given string (Feller has a few pages on it), and facts about a related 'intransitive dice' game (originally from Conway).

A more topical connection is with pattern avoiding permutations, where a finite 'pattern' permutation $\sigma \in \Sigma_k$ is chosen, and a uniform random permutation X is conditioned on 'avoiding' σ, i.e. having no subsequence $i_1 < i_2 < \cdots < i_k$ such that

$$(X(i_1), X(i_2), \ldots, X(i_k)) \text{ is order-isomorphic to } \sigma. \tag{1.3}$$

The same question can be asked for any random sequence X, say iid from a discrete distribution. Do we recover phenomena similar to the permutation case? It seems there is some work on this, in the permutation-avoiding literature, where there are some recursive techniques that apply to general sequences X (not just permutations).

Let X denote a random instance of one of these processes. The over-arching questions we are interested in are:

1. How does the conditioning affect typical properties of X, like the density of each letter of $\mathcal{A}$, or 'random walk' properties of X? **We can compute these kinds of things exactly with linear algebra/generating functions for the limiting SFT.**

2. Is there a simple probabilitistic description of the conditional law X? **The limiting measure is a Markov Chain in the case of SFTs. Gibbs measures give a somewhat nice way to interpolate. Generally speaking X has complex structure.**

3. Viewed as the underlying randomness of a random walk, does a scaled version of X converge to a diffusion? (e.g. if the alphabet is $\{-1, 1\}$, does it converge to BM?) **The book *Analytic Pattern Matching* has some possibly relevant CLTs for this?**

4. For non-trivial sets S, $|\Omega_n(S)| \ll |\Omega_n| = |\mathcal{A}|^n$, so X lives on a set of vanishing measure. Despite this, is there a natural limiting measure as $n \to \infty$, i.e. a measure on Ω_∞ (infinite strings) supported on strings that avoid S? **For 'isomorphic' pattern avoidance, there is a limit in permuton space; for substring avoidance, shift spaces and measures of maximal entropy give a full description.**

5. Is X Markovian, or approximately markovian? Is it possible to construct X one bit at a time by recording the output of a simple markov graph? (Simulations suggest this is possible, up to some 'edge' effects. This would be nice – sampling random pattern avoiding permutations is a hot topic.) **A: For substring avoidance, yes: the measure of maximal entropy on a shift space is a markov chain in some presentation.**

These models seem to have similar flavour to the maximal greedy independent set and the hard core model.

2 Substring patterns

2.1 Definitions

Already the case of excluding single binary words leads to interesting phenomena. Set $\mathcal{A} = \{0, 1\}$, and suppose S is a single word of length l, $S = \{w\}, w = w_1 w_2 \cdots w_l$ for some $w_i \in \mathcal{A}$. The first order of business here is to compute $|\Omega_n(w)|$.

Definition 2.1. *For a fixed word w, let λ_w denote the asymptotic growth rate of $|\Omega_n(w)|$, i.e.*

$$\lambda_w = \lim_{n\to\infty} |\Omega_n(w)|^{1/n}. \tag{2.1}$$

We have that:

Lemma 2.2. *Except in the trivial cases $w = 0, 1, 10,$ or 01, the limit in 2.1 exists and $\lambda_w \in (1, 2)$.*

Proof. λ_w is the topological entropy of the shift of finite type with forbidden word w. □

The entropy λ_w is the Perron-Frobenius eigenvalue of the corresponding edge-shift matrix.

Alternatively, one can compute combinatorially, which involves typical recurison/generating function ideas, but with some novel elements. To compute the count $\Omega_n(w)$, and represent the corresponding process X_n, it helps to construct a corresponding graph.

2.2 Frontier representation

Given a forbidden word w, we construct a directed graph L_w which has state space $\{0, 1, \ldots, |w| = l\}$. This is called the 'follower set edge shift' in symbolic dynamics lingo: infinite paths in the graph L_w will correspond (be in bijection) with the shift space where the word w is forbidden. Each state, aside from state l, (which can, for our purposes, be thought of as a 'graveyard' state with no outgoing edges), has 2 outgoing edges, given by appending both possible letters 0 or 1 to the right end of p, then finding the longest prefix of S that matches a *suffix* of pa (for each $a \in \mathcal{A}$). One of these edges always goes from k to $k+1$. We denote the other edge by

$$d_k = \max\{r : w_1 w_2 \cdots w_r = w_{k-r+2} w_{k-r+3} \cdots w_k \overline{w_{k+1}}\} \tag{2.2}$$

where $\overline{w} = 1 - w$. More generally, if the alphabet is $[q]$ for some positive integer q, state k has q outgoing edges, and exactly one such edge goes from state k to state $k+1$. First, we observe that the set $\mathcal{T}(w)$ of all finite words over $\mathcal{A}$ ending with w is described by paths in the graph L_w in a nice way. To make this map precise, for each edge (k, k') in L_w, let $a(k, k')$ denote its label, i.e. the digit of $\mathcal{A}$ that was appended to the k-prefix of w to obtain the k'-prefix (as a suffix). Also, let $\Gamma(L_S)$ denote the set of all finite paths in L_w starting at state 0 and ending at state l.

Fact 2.3. *The map $f : \Gamma(L_w) \to \mathcal{T}(w)$ given by $f(\gamma) = (a(\gamma_i, \gamma_{i+1}))_i$ (i.e. read the labels of the path) is a bijection.*

Proof. For any finite word v not containing w as a subword, let $m(v)$ denote the length of the maximal suffix of v that is a prefix of w. It suffices to show that

$$\{m(v0), m(v1)\} = \{m(v) + 1, d_{m(v)}\}, \tag{2.3}$$

i.e. that to determine $m(va)$ for some digit a, one only needs to look at the last $m(v)$ digits of v, and that $m(va)$ is given by the endpoint of the corresponding edge in the graph L_w. Observe that for any letter a,$m(va) \leq m(v)+1$, since if va ends with a prefix of w of length j, then v ends with a prefix of w of length $j-1$. In the case where $a = s_{m(v)+1}$, we have $m(va) = m(v)+1$. Otherwise, $m(va) \leq m(v)$, so only the last $m(v)$ digits of v are necessary to determine $m(va)$.

□

These graphs are more special than simply requiring outdegree q everywhere except at state l:

Proposition 2.4. *Let $q = 2$. For any such L_w,*

1. *$d_k \leq k$*

2. *$k - d_k = k' - d_{k'}$ if and only if $k = k'$ or one of $d_k, d_{k'} = 0$.*

Proof. (1) is immediate from the definition. For (2), suppose by contradiction that $k - d_k = k' - d_{k'}$ for some $k \neq k'$ with $k, k' \geq 1$. Then also $d_k \neq d_{k'}$, so assume WLOG $d_k > d_{k'}$. The definition of d_k implies

$$w_{k-d_k+j+1} = w_j \text{ for } j = 1, 2, \ldots, d_k - 1, \text{ and } \overline{w_{k+1}} = w_{d_k}, \tag{2.4}$$

where $\overline{w} = 1 - w$. Taking $j = d_{k'}$ gives

$$w_{d_{k'}} = w_{k-d_k+d_{k'}+1} = w_{k'-d_{k'}+d_{k'}+1} = w_{k'+1}, \tag{2.5}$$

contradicting the last part of 2.4 for k' (as long as neither $d_k, d_{k'} = 0$, in which case one of the equalities is trivial).

□

The above properties are not sufficient to classify all such graphs L_w, since the number of possible sequences of d_k values that satisfy (1) and (2) is strictly larger than 2^{l-1}, while the total number of possible graphs is at most that many (there are 2^l strings, and the bit flip operation $w \to \overline{w}$ preserves the graph, $L_w = L_{\overline{w}}$).

Question 2.5. *Find a full characterization of the sequences of d_k that occur for some string w.*

The graph L_w, viewed as a labeled graph on $\{0, 1, \ldots, l\}$, is enough information to determine the word w, up to permutations of the alphabet (i.e. bit flipping, for the binary alphabet). Let $\mathcal{W}$ denote the family of equivalence classes of words in $[q]^l$ up to bit flip, i.e. $w \sim w'$ if there exists a permutation $\phi : [q] \to [q]$ such that $w_{\phi(i)} = s'_i$ for all $i \in [l]$. Also, let $\hat{\mathcal{L}}$ denote the family of *labeled* directed graphs we obtain from the L_w's, i.e.

$$\hat{\mathcal{L}} = \{L_w : w \in [q]^l\}, \tag{2.6}$$

where the graphs L_w have vertex labels $\{0, 1, \ldots, l\}$ corresponding to the frontier representation given by w.

Lemma 2.6. *There is a (simple, algorithmic) bijection $f : \hat{\mathcal{L}} \to \mathcal{W}$.*

Proof. By the definition of $\hat{\mathcal{L}}$, it suffices to construct an injective $f : \hat{\mathcal{L}} \to \mathcal{W}$. Fix $\hat{L} \in \hat{\mathcal{L}}$, and WLOG assume the image $f(\hat{L})$ has $w_1 = 1$. wuppose we have determined the digits $w_1, w_2, \ldots, w_k$ of $w = f(\hat{L})$. Write $j_1, j_2, \ldots, j_{q-1}$ for the $q-1$ states for which there is an outgoing edge $k \to j_i$ in $\hat{L}$. Note that since these $q-1$ states correspond to appending different letters in $[q]$ to $w_1 \cdots w_k$, then finding the maximal frontier that agrees with a prefix of s, the j_i must all be different. It follows that w_{k+1} is the unique digit in $[q]$ not appearing in the set $\{w_{j_i}, i = 1, 2, \ldots, q-1\}$.

□

It turns out that even if we forget about the vertex labels, no two of the graphs $\hat{L}, \hat{L}' \in \mathcal{L}$ are isomorphic as graphs. Write $\mathcal{L}$ for the same family of graphs as in $\hat{\mathcal{L}}$, but viewed as *unlabeled* directed graphs, and for $\hat{L} \in \hat{\mathcal{L}}$, let L denote the same graph but with labels removed.

Lemma 2.7. *$|\mathcal{L}| = |\hat{\mathcal{L}}|$, i.e. for any $\hat{L}, \hat{L}' \in \hat{\mathcal{L}}$, L and L' are not isomorphic.*

Proof. wuppose $\varphi : \hat{L} \to \hat{L}'$ is an isomorphism of the underlying unlabeled graphs, viewed as a permutation on the labels $[l]$. wince $l \in \hat{L}'$ is the unique state with outdegree 0, we must have $\varphi(l) = l$. Assume by induction that $\varphi(i) = i$ for $i = l, l-1, \dots, j$. By Proposition 2.4 (1), all of the incoming edges to state $j \in \hat{L}'$ has its other end at some state $k \geq j$, except one edge $j-1 \to j \in \hat{L}$. wince $\varphi(j-1)$ must be a state that has an edge $\varphi(j-1) \to j \in \hat{L}'$, and the other values $k \geq j$ are already taken, we must have $\varphi(j-1) = j-1$. wo $\varphi = Id$. □

We also note the following. Let $\mathrm{Rev}(w)$ denote the reversal of the string w, i.e. $\mathrm{Rev}(w) = w_k w_{k-1} \cdots w_2 w_1$.

Lemma 2.8. *$|\Omega_n(w)| = |\Omega_n(Rev(w))|$*

Proof. $\omega \in \Omega_n(w) \iff \mathrm{Rev}(\omega) \in \Omega_n(\mathrm{Rev}(w))$. □

Despite this simple fact, it isn't obvious what the relationship is between L_w and $L_{\mathrm{Rev}(w)}$.

Question 2.9. *Describe a simple mapping $L_w \to L_{Rev(w)}$.*

2.3 Binary example

As an example of the usefulness of the graphs L_w, we work through the necessary computation explicitly for $w = 100$. Here the graph is given by $d_1 = d_2 = 1$. We are trying to solve for $\Omega_n(100)$, which can be thought of as the number of paths in the graph L_{100} of length n, starting at either state 0 or state 1, that never hit state 3. To count these, write $a_n(100)$ as the number of such paths, and partition a_n into three further counts a^0, a^1, and a^2, where a^j is the number of such paths ending at state j. These lead to the following system of recursions, obtained by collecting the incoming edges to each state:

$$a_n^0 = a_{n-1}^0 \tag{2.7}$$
$$a_n^1 = a_{n-1}^0 + a_{n-1}^1 + a_{n-1}^2 = a_{n-1} \tag{2.8}$$
$$a_n^2 = a_{n-1}^1 \tag{2.9}$$

There doesn't seem to be a systematic way to solve such a system, other than plugging in recursively repeatedly until a recursion for a_n appears. In this case, it doesn't take too long:

$$\begin{aligned} a &= a^0 + a^1 + a^2 && (2.10)\\ &= 2a_{-1} - a_{-1}^2 && (2.11)\\ &= 2a_{-1} - a_{-2}^1 && (2.12)\\ &= 2a_{-1} - a_{-3}. && (2.13) \end{aligned}$$

Thus $a_n(100) = 2a_{n-1}(100) - a_{n-3}(100)$, which yields the asymptotic formula

$$a_n(100) \sim \left(1+\frac{2}{\sqrt{5}}\right)\varphi^n, \varphi = \frac{1}{2}(1+\sqrt{5}). \tag{2.14}$$

In general it seems easier to work with the corresponding generating functions $f^j_{100}(z) = \sum_{n\geq 1} a^j_n(100)z^n$ and $f_{100}(z) = \sum_{n\geq 1} a_n z^n$. These functions satisfy $f(z) = f^0(z) + f^1(z) + f^2(z)$ and

$$f^0(z) = z + zf^0(z) \tag{2.15}$$
$$f^1(z) = z + z(f^0(z) + f^1(z) + f^2(z)) \tag{2.16}$$
$$f^2(z) = z + zf^1(z) \tag{2.17}$$

The solution is

$$f^0(z) = \frac{z}{1-z}, f^1(z) = \frac{z}{1-2z+z^3}, f^2(z) = \frac{z^2}{1-2z+z^3}. \tag{2.18}$$

Note that $a^0_n = n$, and asymptotically

$$a^1_n \sim \left(\frac{3+\sqrt{5}}{2\sqrt{5}}\right)\varphi^n, a^2_n \sim a^1_n\varphi^{-1}. \tag{2.19}$$

The proportions of paths that end at $0,1,2$, i.e. $\lim_{n\to\infty}\frac{a^j_n}{a_n}$, are respectively $0, \varphi-1, 2-\varphi$, or $\approx 0, .618, .382$.

wee section 2.5 for a linear algebra approach.

2.4 Letter densities

Computing the average density of 1's is not as simple as the counts $a_n(w)$. Let X_n denote a uniformly random chosen element of $\Omega_n(w)$. In the notation of 2.3,

$$\mathbb{P}(\text{the last digit of } X_n = 1) = \frac{1}{a_n(w)}\sum_{k=1}^{l-1} a^k_n(w)1\{\text{the } k^{th} \text{ digit of } w = 1\}. \tag{2.20}$$

In the example with $w = 100$, we computed $\frac{a^1_n(100)}{a_n(100)} \to \varphi - 1$, so this is the limiting probability of seeing a 1 in the final position. However, this isn't the same as the density of 1's in the whole string, as we will see shortly. The method from 2.3 can likely be extended to compute

$$\lim_{n\to\infty} \mathbb{P}(\text{the } j\text{th digit of } X_n = 1) \tag{2.21}$$

for any fixed $j \in \mathbb{N}$, by enumerating paths in the markov graph L_w 'backwards.' These values should converge, as $j \to \infty$, to the average density of 1's in X_n, γ_w (defined below).

A natural quantity is the density of 1's the string X_n. Consider the average fraction of 1's in a uniformly random w-avoiding string:

Definition 2.10. *For a fixed string w, let γ_w denote the limiting fraction of bits that are 1 over all strings in $\Omega_n(w)$:*

$$\gamma_w = \lim_{n\to\infty} \frac{1}{n|\Omega_n(w)|} \sum_{\omega\in\Omega_n(w)} \#1\text{'s in } \omega. \tag{2.22}$$

How can this density be computed? It seems necessary to further partition the strings $\Omega_n(w)$ into sets $\Omega_{n,k}(w)$, i.e. strings of length n with exactly k 1's. Let $a_{n,k}(w) = |\Omega_{n,k}(w)|$. As an example, we continue with the string $w = 100$. The $a_{n,k}(100)$ satisfy a recursion similar to that for $a_n(100)$, namely

$$a_{n,k} = a_{n-1,k} + a_{n-1,k-1} - a_{n-3,k-1}. \tag{2.23}$$

This can be proved by observing that each $\omega \in \Omega_{n,k}(100)$ can be built from a unique string in $\Omega_{n-1,k}(100) \cup \Omega_{n-1,k-1}(100)$ by appending either a 1 or a 0, except for the ones (of length $n-1$) ending in 10, since adding a 0 would result in a 100. (There is something slightly subtle here. See the definition of *selfless* strings below, and proposition 2.12. 100 is a selfless string.)

wtandard generating function technology yields

$$f(z,w) = \sum_{n,k\geq 0} a_{n,k} z^n w^k = \frac{1}{1 - z(1+w) + z^3 w}, \tag{2.24}$$

and by extracting coefficients and taking limits, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(X_n(i) = 1) = \frac{1}{n a_n(100)}[z^n]\frac{\partial}{\partial w}\Big|_{w=1} f(z,w) \to \frac{5+\sqrt{5}}{10} \approx .7236. \tag{2.25}$$

(As expected, the density of 1's increases as a result of conditioning on avoiding 100.) (Another aside: Mathematica is a bit tempermental about evaluating these kinds of expressions. It seems to be happiest when the derivative in w is evaluated first, then the coefficient of z^n is extracted.) A variance calculation can be performed too:

$$\mathrm{Var}(\text{number of 1's in } X_n) \sim \frac{1}{5\sqrt{5}} n. \tag{2.26}$$

A WLLN follows for the number of 1s, since the variance is $o(n^2)$. (**Note that the number of copies of any string w is asymptotically normal by the k-dependent CLT.**)

Finding these recursions is sometimes very straighforward. In fact, a large class of strings w share common recurrences.

Definition 2.11. *Call a string w **selfless** if no prefix of w matches any suffix of w, i.e. if there exists no $j < l$ such that $w_1 w_2 \cdots w_j = w_{l-j+1} w_{l-j+2} \cdots w_{l-1} w_l$, where w has length l.*

The string $w = 100$ is selfless, and it shares the recurrence above with all other selfless strings of length 3 with a single 1, via the same construction.

Proposition 2.12. *Let w be a selfless string of length l containing exactly j 1's. Then*

$$a_{n,k}(w) = a_{n-1,k}(w) + a_{n-1,k-1}(w) - a_{n-l,k-j}(w). \tag{2.27}$$

Proof. To generate an arbitrary string in $\Omega_{n,k}(w)$, we can start with an arbitrary string of length $n-1$ and append a 0 or a 1. This overcounts things slightly, since adding this final digit may have created an instance of w. wo we need to throw away all strings of length $n-1$ ending with the first $l-1$ digits of w. To complete the proof, it suffices to note the following lemma:

Lemma 2.13. *w is selfless if and only if the map from the set of strings in $\Omega_{n-1,k}(w)$ ending in the first $l-1$ digits of w to $\Omega_{n-l,k-j}$ that chops off the last $l-1$ digits is a bijection.*

$\square$

wince $a_n(w) = \sum_{k=0}^{n} a_{n,k}(w)$, and the 'base case' values $a_{n,k} = \binom{n}{k}$ for $n < l$ or $n = l, k \neq j$ and $a_{l,j} = \binom{l}{j} - 1$ only depend on l and j, we get a large family of stastical coincidences:

Proposition 2.14. *Fix l. If w and w' are any two selfless strings of length l, then $a_{n,k}(w) = a_{n,k}(w')$ and $a_n(w) = a_n w'$ for all n and k. In particular, $\lambda_w = g_{w'}$, and if w and w' have the same number of 1's, then $\gamma_w = \gamma_{w'}$. The common recursion is*

$$a_n(w) = 2a_{n-1}(w) - a_{n-l}(w), \tag{2.28}$$

and λ_w is the unique solution $z \in (1, 2)$ to $z^{l-1} = 1 + z + z^2 + \cdots + z^{l-2}$.

Note that, in contarast to the previous proposition, we don't require that w and w' have the same number of 1's. The only difference is in the base case $n = l$. Solving the recurrence in Proposition 2.12 yields the generating function

$$\sum_{n,k\geq 0} a_{n,k}(w) z^n w^k = \frac{1}{1 - z(1 + w) + z^l w^j}, \tag{2.29}$$

where l is the length of w and j is the number of 1's.

In fact, we have the following more general characterization of λ_w. We now introduce an important object:

Definition 2.15. *Given a word w of length l, its* overlap set $\mathcal{O}$ *is the set*

$$\mathcal{O} = \{i \in [l] : w_1 w_2 \cdots w_i = w_{l-i+1} w_{l-i+2} \cdots w_l\} \tag{2.30}$$

and its correlation polynomial *is the polynomial function*

$$\phi(t) = \sum_{i\in\mathcal{O}(w)} t^{i-1}. \tag{2.31}$$

The (auto-)correlation polynomial $\phi_w(t)$ determines the entropy:

Theorem 2.16. *For two words w, w', $\phi_w(t) = \phi_{w'}(t) \iff \lambda_w = \lambda_{w'}$.*

This is originally due to Guibas and Odlyzdko (1980). The $\implies$ is easy, while the reverse is difficult.

Proof. (Proof of $\implies$) We can give an explicit bijection between $\Omega_w(n)$ and $\Omega_{w'}(n)$ in this case, for every n...? $\square$

Definition 2.17. *Call a string w* **balanced** *if the number of 1's in w is half the length of w.*

Recall $a_n(w) = |\Omega_n(w)|$, the number of strings of length n avoiding w, and $a_{n,k}(w)$ is the number of those with exactly k 1s. We have:

Proposition 2.18. *If w is selfless and balanced, then $\gamma_w = 1/2$. In fact, for all n, the average density of 1's in a uniform random string avoiding w is exactly 1/2, i.e.*

$$\sum_{k=0}^{n} k a_{n,k}(w) = \frac{1}{2} n a_n(w). \tag{2.32}$$

Proof. It would be nice to have a bijective proof. The above can be checked directly using the generating function formula 2.29. wetting

$$f(z,w) = \sum_{n,k\geq 0} a_{n,k}(w) z^n w^k = \frac{1}{1 - z(1+w) + z^l w^{l/2}}, \tag{2.33}$$

and

$$g(z) = \sum_{n\geq 0} a_n(w) z^n = \frac{1}{1 - 2z + z^l}, \tag{2.34}$$

a quick computation shows

$$\left.\frac{\partial}{\partial w}\right|_{w=1} f = \frac{1}{2} z g'(z) \tag{2.35}$$

which is equivalent to the claim. □

Also note: the family of selfless strings is quite large! The probability of a string being selfless is bounded away from 0 for any n (perhaps an interesting computation of its own?), so a constant proportion of strings are selfless. (wimulation suggests the probability of being selfless is approximately .266 for n large. The 'mean field' calculation – i.e. assuming matching each suffix to each prefix are independent events – gives an estimate of $\prod_{j\geq 1} 1 - 2^{-j} \approx .289$.) **There is a recursion for selfless strings which can be solved to some extent. There's an OEIS entry, for example.**

There is another class of strings for which the density can be easily seen to be exactly 1/2. Recall that $\mathrm{Rev}(w)$ is the reversal of w, and $\overline{w_1 w_2 \cdots w_l} = \overline{w_1}\,\overline{w_2} \cdots\ \overline{w_l}$, where $\overline{s} = 1 - s$ is the 'bit flipping' operation. Note that these two operations are commuting involutions, i.e. $\overline{\mathrm{Rev}(w)} = \mathrm{Rev}(\overline{w})$ and $\overline{\overline{w}} = \mathrm{Rev}(\mathrm{Rev}(w)) = w$.

Definition 2.19. *Call a string w **sweet** if $\overline{w} = Rev(w)$.*

Note that sweet strings must be balanced, so all sweet strings have even length. Conditioning on avoiding a sweet string keeps the 0-1 count balanced:

Proposition 2.20. *If w is a sweet string, then $\gamma_w = 1/2$. In fact, for all n, the average density of 1's in a uniform random string avoiding w is exactly 1/2, i.e.*

$$\sum_{k=0}^{n} k a_{n,k}(w) = \frac{1}{2} n a_n(w). \tag{2.36}$$

Proof. It suffices to find a bijection $\omega \mapsto \omega'$ from $\Omega_n(w)$ to itself such that the number of 0's in ω' is equal to the number of 1's in ω. Indeed, the existence of such a bijection implies that the total number of 1's over all strings in $\Omega_n(w)$ is the same as the total number of 0's, which implies the result. The bijection that works has the simple formula $\omega \mapsto \mathrm{Rev} \circ \overline{\omega}$. This map is an involution that swaps 0's and 1's, and that w is sweet implies that it maps $\Omega_n(w)$ to itself. □

It is worth noting that the number of sweet strings grows exponentially, but still makes up a vanishing fraction of all strings. Indeed, the sweet strings of length l can be exactly enumerated by choosing an arbitrary string ω of length $l/2$, then forming the string $\omega \oplus \mathrm{Rev}(\overline{\omega})$, where $\oplus$ is concatenation. So there are exactly $2^{l/2}$ sweet strings of length l.

Being balanced is not enough to guarantee that the conditioned string is balanced. Already there is a counterexample when $l = 4$. Note that of the 6 strings of length 4 with two 1's, up to reversal and bit-flipping only one is not sweet: 1010 and 1100 are sweet, while 1001 is not. And we have:

Fact 2.21. *The limiting density of* 1*'s in a uniform random 1001 avoiding string is*

$$\gamma_{1001} = \frac{2\left(-3+2\sqrt{5}\right)^{5/2}\sqrt{\frac{1}{55}\left(3+2\sqrt{5}\right)}\left(110\sqrt{-3+2\sqrt{5}}+44\sqrt{5\left(-3+2\sqrt{5}\right)}+\sqrt{11}\left(35+17\sqrt{5}\right)\right)}{11\left(-35+27\sqrt{5}\right)\left(\sqrt{11}+3\sqrt{-3+2\sqrt{5}}\right)} \tag{2.37}$$

$$\approx .494161. \tag{2.38}$$

Amazingly, conditioning on avoiding 1001 very slightly decreases the density of 1s!

This ostentatious constant comes from computing with generating functions exactly. (**See section 2.5 for a nicer calculation.**) Via the graph L, one finds recursions (where all a_n are interpreted as $a_n(1001)$ for ease of notation)

$$a_n = 2a_{n-1} - a_{n-3} + a_{n-4} \sim \frac{(27\sqrt{5}-35)(\sqrt{11}+3\sqrt{2\sqrt{5}-3})}{20(2\sqrt{5}-3)^{5/2}\sqrt{3+2\sqrt{5}}}\frac{1}{2^n}(1+\sqrt{3+2\sqrt{5}})^n, \tag{2.39}$$

and

$$a_{n,k} = a_{n-1,k} + a_{n-1,k-1} - a_{n-3,k-1} + a_{n-4,k-1}, \tag{2.40}$$

which satisfies

$$\sum_{n,k\geq 0} a_{n,k} z^n w^k = \frac{1+z^3w}{1-z(1+w)+z^3w-z^4w}. \tag{2.41}$$

Note also that 1001 is not selfless – the recursion for $a_{n,k}$ requires additional 'correction' terms.

2.5 Letter densities via the MME for wFT

Let ν denote the measure of maximal entropy for the shift of finite type with single forbidden word w. This measure can be computed explicitly via some matrix computations with the graph L_w, and gives an alternate way to calculate the entropy λ_w and the letter density γ_w. Namely:

Fact 2.22. λ_w *is the (exponential of the) topological entropy of* ν*, and* γ_w *is* $\nu(C_0)$*, the measure of the cylinder set of* 0 *under* ν*.*

These values can be computed exactly from any representation of the corresponding wFT. λ_w is the largest eigenvalue of any graph representation of the corresponding wFT: L_w is the 'minimal' such representation. As for γ_w, recalling the graph L, and using (by a slight abuse of notation) ν to refer also to the stationary MME – the 'parry measure' – on the graph L, we have:

Proposition 2.23. *Let* $w = w_1 w_2 \dots w_l$ *with* $w_1 = 1$*. Then*

$$\gamma_w = \sum_{k<l:w_k=1} \nu(k).$$

Also, the characteristic polynomial of L_w matches the recursion satisfied by $|\Omega_n|$:

Proposition 2.24. *Let A_w denote the adjacency matrix of L_w, and let $p_w(\lambda)$ denote its characteristic polynomial, say $p_w(\lambda) = \sum_{i=0}^{l} c_i \lambda^i$. Then $a_n = |\Omega_n|$ satisfies*

$$c_0 a_n = \sum_{i=1}^{l} c_i a_{n-i}.$$

Example 2.25. *To illustrate, we recover the example $w = 100$ via this method. The graph L_{100} has adjacency matrix*

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

with characteristic polynomial $\lambda^3 - 2\lambda^2 + 1 = (\lambda - 1)(\lambda^2 - \lambda - 1)$, top eigenvalue $\varphi = \frac{1+\sqrt{5}}{2}$, and right/left eigenvectors

$$r_{100} = \begin{bmatrix} 1 \\ \varphi - 1 \\ 1 - \varphi^{-1} \end{bmatrix} \ell_{100} = \begin{bmatrix} 0 & 1 & \varphi^{-1} \end{bmatrix}.$$

The the parry measure is given by $\nu_j = \frac{1}{Z_{100}} r_j \ell_j, j = 0, 1, 2$, with $Z_{100} = r \cdot \ell$. We have

$$\nu = \frac{1}{3\varphi - 4}(0, \varphi - 1, 2\varphi - 3),$$

and the density of 1s is $\gamma_{100} = \nu(1) = \frac{\varphi-1}{3\varphi-4} = \frac{5+\sqrt{5}}{10}$, since the state where we match the first digit is the only state ending in a 1. This matches the calculation from the previous section.

For completeness we also carry out the analysis this way for:

Example 2.26. *Let $w = 1001$, which has L_{1001} with adjacency matrix*

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

(irreducible over $\mathbb{Q}$) characteristic polynomial $\lambda^4 - 2\lambda^3 + \lambda - 1$, top eigenvalue $\lambda \approx 1.866760$, and right/left eigenvectors

$$r_{1001} = \begin{bmatrix} 1 \\ \lambda - 1 \\ (\lambda - 1)^2 \\ \lambda^{-1} \end{bmatrix} \ell_{1001} = \begin{bmatrix} 1 & \lambda^2(\lambda - 1) & \lambda(\lambda - 1) & \lambda - 1 \end{bmatrix}.$$

Thus the parry measure is

$$\nu = \frac{1}{-2\lambda^3 + 6\lambda^2 - 3\lambda + 3}(1, \lambda^2(\lambda - 1)^2, \lambda(\lambda - 1)^3, 1 - \lambda^{-1}),$$

and $\gamma_{1001} = \nu(1) = \frac{\lambda^2(\lambda-1)^2}{-2\lambda^3+6\lambda^2-3\lambda+3}$.

These exact rational functions for the eigenvectors had to be obtained by hand – so far I don't know a systematic way of determining the exact rational expressions in terms of the top eigenvalue λ.

Question 2.27. *Write a computer program that finds an expression for the Perron eigenvectors as polynomials in the entropy λ.*

We continue with some general computations:

Example 2.28. *Consider $w = 111\cdots 1$, a string of l 1s. The graph $L_{11\cdots 1}$ has adjacency matrix*

$$\begin{bmatrix} 1 & 1 & 0 & \cdots & & \\ 1 & 0 & 1 & 0 & \cdots & \\ 1 & 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & & \\ 1 & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

The characteristic polynomial is $\lambda^n - \lambda^{n-1} - \cdots - \lambda - 1$, with right/left eigenvectors

$$r_j = \sum_{i=1}^{n-j+1} \lambda^{-i}, \ell_j = \lambda^{1-j}, j \in [n].$$

The density of 1*s is $\gamma_{11\cdots 1} = 1 - \nu(0) = \lambda^n(\lambda-1)^2[\lambda^{n+1} - \lambda(n+1) + n]^{-1}$.*

Note the following general fact about the right eigenvector r:

Proposition 2.29. *For any string w, the right eigenvector r_w has strictly decreasing entries. In particular, for any $i = 0, \ldots, |w| - 2$,*

$$(\lambda_w - 1)(r_w)_i > (r_w)_{i+1}. \tag{2.42}$$

Note that since $\lambda_w < 2$, the factor $\lambda - 1 < 1$, so the above is strictly stronger than having decreasing entries: the entries of r decrease at least geometrically at rate $\lambda - 1$.

Proof. We proceed by induction. Since r is a right eigenvector, using $d_i \leq i$ (so $r_{d_i} \geq r_i$ by induction),

$$\lambda r_i = r_{i+1} + r_{d_i} \tag{2.43}$$
$$\geq r_{i+1} + r_i. \tag{2.44}$$

Now subtract. In the base case $i = 0$, we have $d_0 = 0$, so $r_1 = (\lambda - 1) r_0$ is an equality. $\square$

The proof actually shows something a bit stronger: namely that

$$\frac{v_{i+1}}{v_i} \leq \lambda - (\lambda - 1)^{i - d_i}. \tag{2.45}$$

2.6 Extremal entropy

We have the following basic heuristic regarding entropies. Turn the graph L into a markov chain with uniform transition probabilities – so probability $\frac{1}{2}$ on all edges, except $l-1 \to d_{l-1}$ which has probability 1 – and let $\tau_w(i)$ be the hitting time of string w started from state i, with $\mu_w(i) = \mathbb{E}[\tau_w(i)]$, and $\mu = \mu_w = \mu_w(0)$ for short. Then

$$\mu = \sum_{t \geq 0} \mathbb{P}(\tau \geq t) \tag{2.46}$$

$$= \sum_{t \geq 0} \frac{\text{\# paths in } L \text{ started from 0 of length } t}{2^t} \tag{2.47}$$

$$\approx \sum_{t \geq 0} r_w(0)(\lambda_w/2)^t \tag{2.48}$$

$$= r_w(0)\frac{1}{1-\lambda_w/2}. \tag{2.49}$$

This suggests the following:

Theorem 2.30. *$\lambda_w < \lambda_{w'}$ if and only if $\mu_w < \mu_{w'}$.*

This is contained in the Guibas, Odlyzko paper, with a complete characterization (equivalence with zeta function) using the result in Doug Lind's paper. Note that $\mu = \phi_w(2)$, the correlation polynomial evaluated at 2 – this is a classic martingale argument (see section 2.9 for details and a generalization).

The proof in Guibas/Odlyzko is a bit unsatisfying, though! A possible approach to getting the same result for entropy is to work with some poset of all string w (of the same length), with a relation that is easy to work with, and always agrees with $\lambda_w < \lambda_{w'}$. Brian's argument, along with Proposition 2.29, gives

Proposition 2.31. *If $d_i \leq d'_i$ for all i, then $\lambda_w > \lambda_{w'}$.*

(Note also that if $d_i \leq d'_i$ for all i, then $\mu_w > \mu_{w'}$.) A step towards recovering theorem 2.30 along these lines would be:

Conjecture 2.32. *w is selfless if and only if there exists no w' such that $d'_i \geq d_i$ for all i.*

It would suffice to show the 'if' part of this statement, by 2.14.

2.7 Word counts

Rather than jumping straight to the wFT by forbidding a string w, it is natural (and possibly helpful) to study the substring counts $N_w(x) = \#$ of copies of w appearing in $x \in \{0,1\}^n$. Observe that for iid Ber(1/2) bits, as $n \to \infty$,

$$\frac{1}{n}\mathbb{E}N_w \to 2^{-l}, \text{ and } \frac{1}{n}\operatorname{Var}(N_w) \to 2^{-l}(1-2^{-l}) + 2^{-2l}(\mu_w - l), \tag{2.50}$$

where μ_w is the hitting time of w defined above, i.e. the autocorrelation polynomial of w evaluated at 2.

Question 2.33. *Observe that the variance is increasing in μ_w: does this have any significance?*

It is also possible to write an explicit formula for the distribution of N_w, via a matrix calculation. Let X be any SFT where w is an allowable word, and let P be the $\{0,1\}$ valued matrix with 0s at the transitions forbidden by X. Construct matrix Q identical to P, except that we replace any transition $P_{ij} = 1$ which 'creates' a copy of w by a variable y. For example, when X is the shift where $\{11\}$ is forbidden and $w = \{1\}$, we use

$$Q = \begin{bmatrix} 1 & 1 \\ y & 0 \end{bmatrix} \tag{2.51}$$

(Here the states are simply $\{0,1\}$, since the parry measure on the golden mean shift is a markov chain with memory 1.) The number of words of length n containing exactly k copies of w is given by

$$[y^k] \sum_{i,j} (Q^n)_{ij}. \tag{2.52}$$

In the golden mean example, Q has eigenvalues

$$\lambda_{\pm} = \frac{1}{2}\left(1 \pm \sqrt{1+4y}\right), \tag{2.53}$$

and one easily diagonalizes:

$$Q = ADA^{-1}, \text{ where } A = \begin{bmatrix} 1 & 1 \\ \lambda_+ - 1 & \lambda_- - 1 \end{bmatrix}, \ D = \text{diag}(\lambda_+, \lambda_-). \tag{2.54}$$

wome algebra yields that the number of words of length n containing exactly k 1s is

$$[y^k]\frac{1}{\sqrt{1+4y}}\left(\lambda_+^n(1+y) + \lambda_-^n(y - \lambda_-)\right), \tag{2.55}$$

which can be unraveled to get explicit formulas. Computing the joint distribution of (N_w, N_T) for a pair of strings w, T is already a challenge. A similar tool works to get explicit formulas: for example, to count $w = 1$ and $T = 11$ over the full shift (no forbidden words), one would use the matrix

$$\begin{bmatrix} 1 & y \\ y & yz \end{bmatrix} \tag{2.56}$$

with y counting the occurrences of w and z counting the occurrences of T. One could follow the same procedure – diagonalize, then extract coefficients – to obtain some explicit formulas for the N_w and N_T counts. **It's not clear how useful this is. Maybe an appeal to generating function technology can give us a CLT, even a joint CLT?**

2.8 Gibbs measures

There is a natural Gibbs measure that interpolates between iid bits and the wFT obtained by forbidding w: namely, fix $\beta > 0$, and weight occurrences of w by β. That is, for any (fixed) $n \in \mathbb{N}$ we have a measure μ_w on the set of strings of length n given by

$$\mu_{w,\beta}(x) = Z_{w,\beta}^{-1} \exp(\beta N_w(x)), \tag{2.57}$$

where $N_w(x)$ is the number of occurrences of w in x. Then $\beta = 0$ is iid measure, $\beta = -\infty$ is the wFT where w is forbidden, and $\beta = \infty$ is atomic, supported on periodic copies of w. Looking at the observables N_T for $T \neq w$ under the measure $\mu_{w,\beta}$ leads to some interesting questions. For example, $\mathbb{E}_{w,\beta}[N_1]$ is the letter density of 1s in a typical $\mu_{w,\beta}$ sample: if we could show that

$$\frac{\partial}{\partial \beta} \mathbb{E}_{w,\beta}[N_1] < 0 \text{ for } \beta < 0, \tag{2.58}$$

this would imply that $\gamma_w < 1/2$ by integrating over $\beta < 0$. This appears to be true for $w = 1001$, for example.

For many pairs w and T, $\mathbb{E}_{w,\beta}[N_T]$ is either globally minimized or maximized at $\beta = 0$, but this appears to not always be the case.

Question 2.34. *Are $\mathbb{E}_{w,\beta}[N_T]$ and $\mathbb{E}_{T,\beta}[N_w]$ related in a canonical way?*

In general, the derivative has the following nice form:

Proposition 2.35. *Let f be any function on $\{0,1\}^n$. Then*

$$\frac{\partial \mathbb{E}_{w,\beta}[f]}{\partial \beta} = \mathrm{Cov}_{w,\beta}(N_w, f). \tag{2.59}$$

Proof. Note that $Z_{w,\beta} = Z = \sum_x \exp(\beta N_w)$, so

$$\frac{\partial Z}{\partial \beta} = Z \mathbb{E}[N_w]. \tag{2.60}$$

Thus

$$\begin{aligned} \frac{\partial}{\partial \beta} \mathbb{E}[f] &= -Z^{-2} \mathbb{E}[N_w] \sum_x f(x)\mu(x) + Z^{-1} \sum_x f(x) N_w(x) \mu(x) && (2.61) \\ &= \mathbb{E}[N_w \cdot f] - \mathbb{E}[N_w] \cdot \mathbb{E}[f] && (2.62) \\ &= \mathrm{Cov}(N_w, f). && (2.63) \end{aligned}$$

□

In particular:

Corollary 2.36. *At $\beta = 0$, $\frac{1}{\partial \beta} \partial \mathbb{E}_{w,\beta}[N_T] = \frac{1}{\partial \beta} \partial \mathbb{E}_{T,\beta}[N_w]$.*

Here is an attempt to calculate $\mathbb{E}_{w,\beta}[N_w]$ for all β, which seems to nearly work(!) Recall that μ is a measure on $\{0,1\}^n$ for fixed (large) n. Fix w, and write

$$g_k(\beta) = n^{-k} \mathbb{E}_{w,\beta}[N_w^k], \tag{2.64}$$

the kth moment of $n^{-1}N_w$ under $\mu_{w,\beta}$. By Proposition 2.35,

$$\frac{\partial g_k(\beta)}{\partial \beta} = n(g_{k+1}(\beta) - g_k(\beta)g_1(\beta)). \tag{2.65}$$

Write $m_\beta = \lim_{n\to\infty} g_1(\beta)$ for short, and $\sigma_\beta^2 = \lim_{n\to\infty} \operatorname{Var}(n^{-1}N) = \lim_{n\to\infty} n^{-1}\left(\mathbb{E}[N^2] - n^2c_\beta^2\right)$. At $\beta = 0$, N is an l-dependent sum of random variables, so it satisfies a CLT, i.e. viewed as a random variable in the measure $\mu_{\beta,w}$ for fixed $\beta = 0$ and $n \to \infty$,

$$\frac{N - g_1 n}{\sigma\sqrt{n}} \to_d N(0,1). \tag{2.66}$$

This should also be true for any $\beta \in \mathbb{R}$, but it requires understanding the infinite-volume limit of the measures μ better. (Proof coming soon?) Using this as an approximation, we have that for each fixed β,

$$N \approx m_\beta n + \sigma_\beta\sqrt{n}Z, \tag{2.67}$$

where Z is a Normal$(0,1)$. This gives approximations to the moments of N:

$$\begin{aligned}
\mathbb{E}[N^k] &\approx \mathbb{E}[(m_\beta n + \sigma_\beta\sqrt{n}Z)^k)] &(2.68)\\
&= m_\beta^k n^k + k m_\beta^{k-1}\sigma_\beta n^{k-1/2}\mathbb{E}[Z] + \binom{k}{2} m_\beta^{k-2}\sigma_\beta^2 n^{k-1}\mathbb{E}[Z^2] + \cdots &(2.69)\\
&= m_\beta^k n^k + \binom{k}{2} m_\beta^{k-2}\sigma^2 n^{k-1} + O(n^{k-2}). &(2.70)
\end{aligned}$$

In particular, $\mathbb{E}[N] \approx m_\beta n$,

$$\mathbb{E}[N^2] \approx m_\beta^2 n^2 + \sigma_\beta^2 n, \tag{2.71}$$

and

$$\mathbb{E}[N^3] \approx m_\beta^3 n^3 + 3m_\beta\sigma_\beta^2 n^2. \tag{2.72}$$

Now we plug into Proposition 2.35:

$$\frac{\partial m_\beta}{\partial \beta} \approx \lim_{n\to\infty} \frac{1}{n}\operatorname{Cov}(N,N) = \sigma_\beta^2, \tag{2.73}$$

and

$$\begin{aligned}
\frac{\partial \sigma_\beta^2}{\partial \beta} &= \lim_{n\to\infty} \frac{\partial}{\partial\beta} n^{-1}\operatorname{Var}(N) &(2.74)\\
&= \lim_{n\to\infty} n^{-1}\left(\frac{\partial}{\partial\beta}\mathbb{E}[N^2] - \frac{\partial}{\partial\beta}(\mathbb{E}[N]^2)\right) &(2.75)\\
&= \lim_{n\to\infty} n^{-1}\left(\mathbb{E}[N^3] - \mathbb{E}[N^2]\mathbb{E}[N] - 2\mathbb{E}[N]\operatorname{Var}(N)\right) &(2.76)\\
&\approx \lim_{n\to\infty} n^{-1}(m_\beta^3 n^3 + 3m_\beta\sigma_\beta^2 n^2 - (m_\beta^2 n^2 + n\sigma_\beta^2)(m_\beta n) - 2m_\beta\sigma_\beta^2 n^2) &(2.77)\\
&= 0 \quad (!!!) &(2.78)
\end{aligned}$$

(Here we used Proposition 2.35 twice, plus the product rule, to evaluate the derivatives.)

What happened? The expression for the derivative of σ_β^2 cancelled to constant order, so everything vanished! I believe most of this sketchy calculation is correct, except that the normal approximation to N is not good enough: with a better approximation, we would see that this expression *does not cancel* to linear order $\Theta(n)$, so we would get some actual expression in the limit.

Question 2.37. *Maybe the combinatorial/Markovian structure of the measure μ is enough to compute $\mathbb{E}[N], \mathbb{E}[N^2]$ and $\mathbb{E}[N^3]$ precisely enough so that this calculation can be carried out?*

If it works, this would be a kind of bootstrapping: we first get good enough *approximations* to the moments of N so that we can solve for the moments explicitly via a system of differential equations! It may be that further moments of N pop out, so it may be a system involving the first l moments of N, or perhaps the full moment generating function would be required.

One can also jump directly to the moment generating function:

$$g(\beta, w) = \sum_{k \geq 0} g_k(\beta) w^k. \tag{2.79}$$

Playing with this a bit gives a functional relation on g:

$$\frac{\partial}{\partial \beta} g(\beta, w) = 1 + n \sum_{k \geq 1} w^{-1} w^{k+1} g_{k+1}(\beta) - n g_1(\beta) w^{-1} \sum_{k \geq 1} g_k(\beta) w^k \tag{2.80}$$

$$= 1 + n w^{-1}(g(\beta, w) - 1 - w g_1(\beta)) - n g_1(\beta)(g(\beta) - 1) \tag{2.81}$$

$$= 1 - n w^{-1} + n(w^{-1} - g_1(\beta)) g(\beta, w). \tag{2.82}$$

We also know the 'initial values' $g(\beta, 0) = 1$, and $g(0, w)$ can be calculated directly (see the 'Analytic Pattern Matching' book for explicit formulas. $\beta = 0$ is the iid case.) Finally,

$$g_1(\beta) = \frac{\partial}{\partial w}\Big|_{w=0} g(\beta, w). \tag{2.83}$$

wo g is described by the single PDE

$$g_\beta(\beta, w) = 1 - n w^{-1} + n g(\beta, w)(w^{-1} - g_w(\beta, 0)). \tag{2.84}$$

Again we have the same 'normalization' issue – if we understood g better, the factor of n would cancel out and we would be left with a proper, parameter-free PDE. But this requires getting better approximations for the moments of N.

Question 2.38. *Does the function $\mathbb{E}_{w,\beta}[N_T]$ always have a single critical point? When is it a global max/min? When is the extreme point at $\beta = 0$?*

wome computer simulations have been carried out for this. I used Glauber dynamics to approximate the function $f_{w,T}(\beta) = \mathbb{E}_{w,\beta}[N_T]$ for large n, and $\beta \in [-b, b]$ for $b \approx 5$. The function f always appears to be smooth (probably analytic), the limit $\beta \to -\infty$ exists and agrees with γ_w, and the limit $\beta \to \infty$ exists and agrees with the density of T's in the 'periodic tiling of $\mathbb{Z}$ by w's. Here are the results I recorded:

- $T = 1$

  - $w = 1001$: f has a maximum at $\beta = 0$
  - $w = 100$: f strictly decreasing
  - $w = 11$: f strictly increasing
  - $w = 100110$: f always exactly 1/2 (clear by symmetry)
  - $w = 110010$: f always exactly 1/2 (not sure why??)
  - $w = 10011100$: f always exactly 1/2 (not sure why??)

- $T = 11$
  - $w = 00$: f strictly decreasing
  - $w = 1001$: f strictly decreasing
  - $w = 111$: f strictly increasing
  - $w = 10011100$: f strictly increasing

- $T = 00$, $w = 1001$: f decreasing then increasing, minimum around $\beta = 1.2$ (!!)

- $T = 10$, $w = 1001$: f strictly increasing

- $T = 01$, $w = 1001$: f strictly increasing

- $T = 0000$, $w = 1001$: f strictly decreasing

- $T = 11, w = 101$: f has a maximum at $\beta = 0$

I also simulated $\mathbb{E}_{w,\beta}[N_w]$ as a function of β: it is strictly increasing, and appears to depend only on the correlation polynomial of w, i.e. these functions are identical for w, w' with the same correlation polynomial. (Perhaps this can be shown directly?)

2.9 Hitting times

Suppose we have a process $X_1, X_2, \ldots$ with the natural filtration $\mathcal{F}_t = \sigma[(X_s)_{s\le t}]$, and X_t taking values in our (finite) alphabet $[q]$. We're thinking of the X_i as digits generated from the MME of some SFT – e.g. iid digits in the case of the full shift – so in our applications the X sequence has finite memory, but may not be Markov (with memory 1). (The upcoming martingale construction is general though, and doesn't require this.) Fix a word w of length l, and consider the stopping time $T_w = \min\{t : (X_{t-l+1}, X_{t-l+2}, \ldots, X_t) = w\}$. In this section we derive a general formula for $\mu_w = \mathbb{E}[T_w]$, which in the case that X is actually a Markov chain reduces to a relatively simple formula closely related to the autocorrelation polynomial of the word w.

The plan is to build a martingale involving the word w, which is described by a certain gambling game (betting on occurrances of the string w), and apply the optional stopping theorem at time T. For any $t \ge 0$, any finite string $x \in [q]^t$, and any $i \in \{1, 2, \ldots, k\}$, say the triple (i, t, x) is *streaking* if the last i letters of x are the first i letters of w (any triple with $i = 0$ is streaking), and define

$$Q(i, t, x) = \mathbb{P}(X_{t+1} = w_{i+1} | (X_1, \ldots, X_t) = x),$$

and if $Q(i, t, x) \in (0, 1)$ let

$$G(i, t, x) = \begin{cases} \prod_{j=0}^{i-1} Q(j, t-i+j, (x_1, \ldots, x_{t-i+j}))^{-1}, & (i, t, x) \text{ streaking} \\ 0, & (i, t, x) \text{ not streaking} \end{cases}$$

(else if $Q(i, t, x) \in \{0, 1\}$ then $G(i, t, x) = G(i-1, t-1, (x_1, \ldots, x_{t-1}))$. Also denote by $\mathrm{Bet}(i, t, x)$ the probability distribution taking value $G(i, t, x)Q(i, t, x)^{-1}(1-Q(i, t, x))$ with probability $Q(i, t, x)$ and value $-G(i, t, x)$ with the complementary probability if $Q(i, t, x) \notin \{0, 1\}$ and (i, t, x) is streaking, and Bet $= 0$ deterministically otherwise. To explain the terms, imagine that a bettor arrives at each time $t \ge 0$ and bets one dollar on the digits of w occurring in X in order, started from digit t, and reinvests all her winnings on the next digit if she wins, or goes home if she ever loses. Then a triple (i, t, X) is streaking if the bettor who arrived at time $t - i + 1$ bets on the correct digit at least i times, $Q(i, t, X)$ is her probability of betting correctly on the $i + 1$st digit, and $G(i, t, X)$ is her total fortune (including the initial 1 dollar investment) up to the ith digit, and $\mathrm{Bet}(i, t, X)$ is her gamble on the $i + 1$st digit. The bet amounts are arranged so that each bet is fair (mean zero): if some amount g is bet on a digit that has probability q to occur, then our net gain is $gq^{-1} - g$ if we win (with probability q) and $-g$ if we lose (probability $1 - q$), which has expected value

$$q \cdot (q^{-1}g - g) + (1 - q) \cdot (-g) = 0.$$

We use all these bets to define a martingale W_t which is the total *net* fortunes of all bettors up to time t, given by

$$W_t = -t + \sum_{i=1}^{l} G(i, t, (X_s)_{s\le t}).$$

Observe that conditionally on $(X_s)_{s\le t} = x$, the increment $\Delta W_t = W_{t+1} - W_t$ is equal in distribution to

$$\sum_{i=0}^{l-1} \mathrm{Bet}(i, t, x),$$

which has expectation zero (by linearity of expectation). It follows that $\mathbb{E}[W_{t+1}|\mathcal{F}_t] = W_t$, i.e. W is indeed a martingale with $\mathbb{E}W_t = 0$ for all $t \geq 0$. In the case where the underlying sequence X arises from the MME of an irreducible SFT, T_w is sub-geometrically distributed, i.e.

$$\mathbb{P}(T_w > t) \leq \exp(-ct)$$

for some $c > 0$, and thus T and W_t satisfy the conditions of the optional stopping theorem (OST). (This is an easy general fact about irreducible markov chains: hitting times are always sub-geometric. So since X is markov in some higher block representation, we also get exponential decay in the lower block representation, with constant c scaled by the ratio between the block lengths.) Applying the OST at time T, so all bets on digits up to $X_T = w_l$ have been settled, yields

$$\mathbb{E}W_{T_w} = -\mathbb{E}T + \sum_{i=1}^{l} \mathbb{E}[G(i, T, (X_s)_{s\leq T})] = \mathbb{E}W_0 = 0.$$

Looking back at the definition of G, the ith term in this sum is nonzero exactly when the last i digits of w are equal to the first i digits of w, since at time T, the last l digits of X are w. Thus we can write

$$\mathbb{E}T = \sum_{i\in\mathcal{O}(w)} \mathbb{E}[G(i, T, (X_s)_{s\leq T})],$$

where $\mathcal{O}(w)$ is the overlap set of w (i.e. the set of i such that the first i digits of w match the last i digits of w.) Finally, in the case where X is itself a maximal entropy markov chain arising from an wFT, the values $Q(i, t, x)$ appearing in the product $G(i, t, x)$ depend only on the last digit of x, which for $i < l$ and $t = T$ is the deterministic digit w_{l-i}, and have the form

$$\begin{aligned} G(i, T, (X_s)_{s\leq T}) &= \left(\prod_{j=1}^{i} \mathbb{P}(X_1 = w_{l-j+1}|X_0 = w_{l-j})\right)^{-1} \quad (\text{for } i \in \mathcal{O}(w) \setminus \{l\}) \\ &= \exp(\lambda i) r(w_{l-i}) r(w_l)^{-1}, \end{aligned}$$

(where we used the fact that $i \in \mathcal{O}(w)$ to get $w_i = w_l$) and the one special case

$$G(l, T, (X_s)_{s\leq T}) = \exp(\lambda l) r(X_{T-l}) r(w_l)^{-1},$$

where λ is the entropy of the MME and ℓ and r are the left and right probability eigenvectors (with eigenvalue λ) of the edge shift graph on $[q]$ representing the SFT X, scaled so that ℓ is a probability vector and $\ell^T r = 1$. Putting this together, we get:

Theorem 2.39. *Let X be a markov chain that realizes a measure of maximal entropy for a shift of finite type, and let w be a finite word (of length l) in the language of X. Then the hitting time T_w of the word w in X satisfies*

$$\mathbb{E}[T_w] = r(w_l)^{-1} \left(e^{\lambda l} \mathbb{E}[r(X_{T-l})] + \sum_{i\in\mathcal{O}(w)\setminus l} e^{\lambda i} r(w_{l-i}) \right) \tag{2.85}$$

Example 2.40. *When X is the full shift over $[q]$, i.e. X is iid over $[q]$, $\lambda = \log q$, $\ell = q^{-1}\vec{1}$ and $r = \vec{1}$, and we get*

$$\mathbb{E}[T_w] = \sum_{i \in \mathcal{O}(w)} q^i = q\phi_w(q)$$

where $\phi_w(q)$ is the auto-correlation polynomial of w.

More generally, whenever r is a constant vector, we don't have to deal with the pesky expectation in Theorem 2.39.

Example 2.41. *Suppose the edge shift of X has uniform in-degree, i.e. the edge shift matrix for X is doubly stochastic (and let λ denote the entropy). Then $r = \vec{1}$ is the right eigenvector for eigenvalue λ, so we obtain the same formula:*

$$\mathbb{E}[T_w] = e^{\lambda}\phi_w(e^{\lambda}). \tag{2.86}$$

An example of such a shift is with $q = 3$, and the forbidden words $\mathcal{F} = \{11, 22, 33\}$. Then the 1-block representation is a markov chain with edge shift matrix $J_3 - I_3$, i.e. the matrix of all 1s except for 0's on the diagonal, and we have $\lambda = \log 2$ and right eigenvector $\vec{1}$.

Example 2.42. *Let X be the golden mean shift, i.e. with forbidden word $\{11\}$ over alphabet $\{0, 1\}$, and assume $w_1 = 1$. One computes directly that, for the 2 by 2 matrix representation of X, with entropy $\log \varphi = \log \frac{1+\sqrt{5}}{2}$, $\nu(0, 0) = \varphi^{-1}, \nu(0, 1) = \varphi^{-2}$. Thus for $x = (x_0, x_1, \ldots, x_k)$,*

$$\mathbb{P}(X = x | X_0 = x_0) = \varphi^{-N_{00}(x)}\varphi^{-2N_{01}(x)},$$

where $N_{00}(x)$ and $N_{01}(x)$ count the number of 00 or 01 subwords of x, respectively. But it's an easy exercise (by induction, for example) that for any x with initial and final digits x_i and x_f,

$$N_{00}(x) + 2N_{01}(x) = len(x) - 1 - x_i + x_f \tag{2.87}$$

wo using the assumption $w_1 = 1$, which implies $X_{T-l} = 0$ deterministically, (and also $w_{l-i} = 0$ for $i \in \mathcal{O}(w)$, since for such i $w_{l-i+1} = w_1 = 1$)

$$\mathbb{E}[T_w] = \varphi^{1+w_l}\phi_w(\varphi) \tag{2.88}$$

Question 2.43. *Compute explicitly some other small example that doesn't fit into any of the above examples. Do we still get a similar formula?*

2.10 Conjectures/questions

Some collected conjectures/questions from the above:

Question 2.44. *To what extent do the Guibas/Odlyzko results hold over subshifts?*

Here is our best guess as of October 2023:

Conjecture 2.45. *Let X be an irreducible shift of finite type with entropy λ, and let a, b be allowable words in X. Let λ^a and λ^b denote the entropies of the further subshifts obtained by additionally forbidding the words a or b. Then $\lambda^a \leq \lambda^b \iff \phi_a(\lambda) \leq \phi_b(\lambda)$, where ϕ is the auto-correlation polynomial.*

Conjecture 2.46. *The density of 1's in a uniform random element of $\Omega_n(w)$ is 1/2 if and only if w is sweet or balanced and selfless and satisfies ...*

Simulation found counterexamples of length 8 to just being sweet/balanced, namely 10011010 and 10100110.

Conjecture 2.47. *$\gamma_w = 1/2$ only if w is balanced.*

This has been confirmed by (approximate) simulations up to strings w of length 20.

Conjecture 2.48. *$\gamma_w > 1/2$ if and only if w has at least as many 1's as 0's.*

Question 2.49. *Classify the set of balanced strings w with $\gamma_w < 1/2$.*

Conjecture 2.50. *For all l and all strings w of length l, $|\gamma_w - 1/2| \leq C\exp(-cl)$. If this holds, what is the optimal rate c?*

Guiding questions moving forward:

Question 2.51. *For any shift space, describe the strings / forbidden sets that give maximum/minimum entropy.*

Question 2.52. *Carry out the same analysis with topological pressure, i.e. for underlying measure Ber(p) for arbitrary $p \in (0,1)$ instead of just $p = 1/2$. How do the densities/entropies depend on p?*

REX project (UBC undergrads) looked Penny's game with arbitrary p. We found a few things, including: when $p = 1/2$, the only time a longer string beats a shorter string, i.e. probability of appearing first in an iid Ber(p) sequence is $> 1/2$, is when the shorter string has auto-correlations of all lengths; and in the limit $p \to 0$, for fixed strings v, w, v beats w with probability in $\{0, 1/2, 1\}$.

3 Subsequence patterns in iid sequences

Fix a 'pattern' of length k, i.e. a $\sigma \in [k]^k$, and let Z_n be iid according to some fixed, discrete distribution p on $\mathbb{N}$, i.e. $\mathbb{P}(Z = j) = p_j$ for $j = 1, 2, \ldots$. Let X_n^σ be the conditional measure of $(Z_1, Z_2, \ldots, Z_n)$ on avoiding σ as a sub-pattern, in the usual sense (see the introduction). Note that σ can have repeated elements. For example, if $\sigma = 112$, then $(Z) = 13222$ is σ avoiding, but $(Z) = 13223$ is not. What can we say about X? As a first example, consider:

3.1 $\sigma = 11$, arbitrary distribution

This is equivalent to conditioning that $Z_1, \ldots, Z_n$ are distinct. For an arbitrary distribution p, we have the formula

$$\mathbb{P}((Z)_n \text{ is } 11-\text{avoiding}) = n! \sum_{|A|=n} \prod_{a \in A} p_a = n! E^n(p), \tag{3.1}$$

where the sum is over all subsets of $\mathbb{N}$ of size a. This is known to combinatorialists as ($n!$ times) the 'elementary homogeneous symmetric polynomial,' over the variables $p_1, p_2, \ldots$. We can also write inclusion probabilities in this way:

$$\mathbb{P}(j \in X_n) = \frac{1}{E^n(p)} \sum_{j \in A, |A|=n} \prod_{a \in A} p_a = p_j \cdot \frac{E^{n-1}(p_{\setminus j})}{E^n(p)}, \tag{3.2}$$

where $p_{\setminus j}$ denotes the sequence of p_i's, but with p_j removed. More formulas can be obtained like this, but it's not clear what they're useful for.

3.2 Uniform distribution

A natural setting is to take Z to be a uniform random variable on $[N]$ for some large integer N, and take n to be some function of N. Note that if N is much larger than n, say $n = \log N$, then it's nearly identical to the situation where Z is uniform on $(0, 1)$, which is *exactly* the case of pattern avoiding uniformly random permutations.

So think of n as being large enough compared to N that there is a non-vanishing probability of choosing the same element twice, i.e. when $N = O(n^2)$. Let $\mathcal{A}(\sigma, n, N)$ denote the set of σ-avoiding strings of length n over the alphabet $[N]$. For example, $X_{n,N}^{11}$ is simply a uniform random subset of $[N]$ of size n, and $|\mathcal{A}(11, n, N)| = \binom{N}{n}$. More interesting is X^{12}, i.e. conditioning Z to be non-increasing. These are not too hard to count:

$$|\mathcal{A}(12, n, N)| = \binom{N + n - 1}{n} \tag{3.3}$$

by a typical 'stars and bars' count. Note that $X_{n,N}^{12}$ can be thought of as a uniformly random element of $\mathcal{A}(12, n, N)$, since there is a unique order of the elements of X making it non-increasing. An interesting quantity to study here is $M_{n,N} = \max X_{n,N}$. Some calculations with binomials yield that:

Lemma 3.1. *Fix $\lambda > 0$. As $N \to \infty$, we have the distributional convergence*

$$N - M_{\lfloor \lambda n \rfloor, N} \to Geo(1 + \lambda), \tag{3.4}$$

i.e.

$$\mathbb{P}(M_{\lfloor \lambda n \rfloor, N} = N - s) \to \lambda(1+\lambda)^{-1-s} \textit{ for } s = 0, 1, 2, \ldots. \tag{3.5}$$

Thus the maximum value of X is tight to N for $n = O(N)$, and the distance away from N is geometrically distributed, with parameter $1 + \lambda = 1 + n/N$. For example, when $n = N, \lambda = 2$, so the maximum is $\mathrm{Geo}(1/2)$ away from N.

Question 3.2. *Come up with a simple combinatorial explanation for this phenomenon.*

Todo: figure out how it works for $n = \sqrt{N}$ or $n = N^\beta$. There should be a similar limit theorem with some geometric/exponeitially distributed distance. For example, when $n = \sqrt{N}$, the distance should be on order $\sqrt{N}$, I think – after scaling properly, what do we get?