ON THE EVOLUTION OF RANDOM STRINGS. A SHORT SURVEY

SERGEY DOVGAL

Abstract. A random walk on words is a random process that takes at most d letters at the end of a string and replaces it with at most d letters. A walk is called ergodic if the probability to return to any given state tends to a positive value. I will describe the results and the techniques that allow to characterise ergodicity for $d = 1$ using generating functions, and for larger d with probabilistic techniques, obtained by Gajrat, Malyshev and Menshikov in 1993.

COFINANCÉ PAR L'UNION EUROPÉENNE
RÉGION BOURGOGNE FRANCHE COMTE

Contents

1. Introduction

First of all, let us define a Markov Chain on strings.

Definition 1. A random walk on strings is a sequence of random words X_n on the set of all finite words on a finite alphabet Σ whose transitions satisfy the given properties:

- At most d letters are removed from the end of the string and replaced by at most d letters.
- The transition probabilities depend on the top d letters of the string, provided that its length is at least d, or, on the whole string, if its length is strictly less than d.

Markov Chain processes on strings can be viewed from many different angles. Let us name a few.

(1) **Random walks on regular languages**. The most classical type of a random walk that people consider is a walk on integer points in multiple dimensions, possibly restricting to positive half-hyperplanes. However,

walks on other infinite structures receive a lot of attention as well. The typical instances are walks on infinite groups, such as $PSL(2,\mathbb{Z})$. On the other hand, the criterion of "positivity" can be fulfilled if we are walking on a monoid, where an inverse operation is not always defined. Without a positivity condition it is difficult to imagine random walks that can return to zero with a positive probability.

(2) **Queueing theory**. Since the evolution of such strings is more or less equivalent to stack operations, this corresponds to an LIFO queue and operations on it.

(3) **Networks**. Normally, a walk $(X_n)_{n\geqslant 0}$ is parametrised by an integer number n which represents discrete time. In more complicated situations, we can imagine that the underlying parameter n is a node of some graph or lattice. Then, the stack will represent nested packets, or consecutive applications of an encryption/decryption algorithm. In this model, it is natural to ask whether a package can arrive from the source to the destination.

(4) **A pushdown automaton** is a natural device to work with evolutions of random strings. In a more general version, it gives rise to *non-deterministic walks* on strings which is an even more complicated model[1].

The probability to return to zero is a function of the transition probabilities. We will describe the region in the space of transition probabilities in which the probability to return to zero is positive in the limit.

2. The case $d = 1$.

Suppose that we modify at most one letter at a time. We will show that this model leads to a very simple condition of ergodicity.

The transition probabilities of a random walk satisfy

$$\begin{aligned}
\mathbb{P}(X_{n+1} = x_1x_2\dots x_m a'b \mid X_n = x_1x_2\dots x_m a) &= p(a'b|a),\\
\mathbb{P}(X_{n+1} = x_1x_2\dots x_m a' \mid X_n = x_1x_2\dots x_m a) &= p(a'|a),\\
\mathbb{P}(X_{n+1} = x_1x_2\dots x_m \mid X_n = x_1x_2\dots x_m a) &= p(\varnothing|a),\\
\mathbb{P}(X_{n+1} = a \mid X_n = \varnothing) &= p(a|\varnothing),\\
\mathbb{P}(X_{n+1} = \varnothing \mid X_n = \varnothing) &= p(\varnothing|\varnothing).
\end{aligned} \tag{1}$$

Definition 2. Green's function. The following function is introduced in Lalley's paper[2].

$$G_{xy}(z) = \sum_{n=0}^{\infty} \mathbb{P}(X_n = y \mid X_0 = x) z^n.$$

Under certain aperiodicity assumptions, all the functions $G_{xy}(z)$ have the same radius of convergence.

Proof idea. Later, we obtain a system of equations defining these functions $G_{xy}(z)$. The system is strongly connected and we can truncate it to a finitely many

[1] Elie De Panafieu, Mohamed Lamine Lamali, Michael Wallner. Combinatorics of nondeterministic walks of the Dyck and Motzkin type.

[2] Stephen P. Lalley. Random Walks on Regular Languages and Algebraic Systems of Generating Functions

short strings x and y. In this case, we can only apply the implicit function theorem as long as the derivative of the defining system does not turn to zero.

Implicit function theorem, Example. Consider the equation of a circle:

$$x^2 + y^2 = R^2.$$

It can define a function $y = y(x)$, but there are two branching points $x = \pm R$, where the function is not anymore locally linear, but has the form

$$y = \pm\sqrt{R^2 - x^2}.$$

Let us rewrite the initial equation in the form

$$f(x, y) = 0, \quad f(x, y) = x^2 + y^2 - R^2.$$

The conditions of the Implicit Function theorem require that $\partial_y f(x, y) \neq 0$. This condition is violated if and only if $y = 0$ which implies $x = \pm R$.

When the condition is violated, then $x(y)$ is typically *locally quadratic* (due to the choice of the next term in the Taylor expansion), and this corresponds to a square-root type singularity.

Note that at each of the points $x \in \{-R, R\}$, we will have two different local expansions

$$y(x) = \pm\sqrt{R^2 - x^2} \sim_{x\sim -R} \pm\sqrt{2}R \cdot \sqrt{1 + \frac{x}{R}}$$

and

$$y(x) = \pm\sqrt{R^2 - x^2} \sim_{x\sim R} \pm\sqrt{2}R \cdot \sqrt{1 - \frac{x}{R}}.$$

We might also remove the $\pm$ in front of the square root to denote the complex-valued square root with two branches.

In some very specific cases, we have more complex scenarios than just the square-root type singularity, but this is out of the scope of this survey.

The idea behind this corner-case of the implicit function theorem is to prove that all the unknown functions of the system of polynomial equations share the same singular point ρ and have an asymptotic representation

$$F(z) = a - b\sqrt{1 - \frac{z}{\rho}} + \mathcal{O}\left(\left|1 - \frac{z}{\rho}\right|\right).$$

around this point as $z \to \rho$.

Proposition 1. If the Markov Chain is ergodic, then the radius of convergence of $G_{xy}(z)$ is equal to 1. Moreover, in this case, $G_{xy}(1) = \infty$.

Proof. Use the ratio test for series convergence, and the fact that the sequence tends to a constant. □

This criterion will be used in the other direction as well, but let us abstain from formulating rigorous condition for this to be sufficient.

2.1. **Geometric view.** Recall that we are interested in identifying the region where the Markov Chain is ergodic.

To do that, we need a family of functions $H_a(z)$, where a is from the alphabet. Let, by definition, $H_a(z)$ be the generating function of walks that start with a letter a and finish with an empty word, with addition condition that an empty word does not appear before.

Formally speaking,

$$H_a(z) := \sum_{n=1}^{\infty} z^n \mathbb{P}(X_n = \varnothing \mid X_0 = a, |X_i| \geqslant 1 \text{ for } i = 0, 1, \ldots, n-1).$$

If $|x| \leqslant 1$ and $|y| \leqslant 1$, we can write

$$G_{xy}(z) = \mathbf{1}_{[x=y]} + z \sum_{a\colon |a| \leqslant 1} p(a|x) G_{wy}(z) + z \sum_{ab} p(ab|x) H_b(z) G_{ay}(z).$$

This is a system of linear equations. It admits a matrix form:

$$\mathbf{G}(z) = \mathbf{I} + z\mathbf{P}\mathbf{G}(z) + z\mathbf{K}(z)\mathbf{G}(z),$$

where $\mathbf{P}$ is the matrix of transition probabilities $\mathbf{P} = (p(y|x))_{y,x}$, and

$$\mathbf{K}(z) = \left(\begin{cases} \sum_b p(yb|x) H_b(z), & \text{if } |y| = 1 \\ 0, & \text{if } |y| = 0 \end{cases} \right)_{xy}.$$

This yields, for sufficiently small modulus $|z|$,

$$\mathbf{G}(z) = (\mathbf{I} - z\mathbf{P} - z\mathbf{K}(z))^{-1}.$$

Now, let us find the defining system for the functions $H_a(z)$. By running over all possible initial jumps, we obtain a system of quadratic equations

$$H_a(z) = zp(\varnothing|a) + z \sum_x p(x|a) H_x(z) + z \sum_{xy} p(xy|a) H_y(z) H_x(z).$$

Note that in general, it is far from being possible to present an explicit solution to a system of quadratic equations, because they lead to an equation whose degree is more than 4. However, in this particular case there is a funny trick.

Since we are looking into ergodicity, we are looking at the values of the generating functions at $z = 1$. The above system has a *trivial* solution $H_a(1) = 1$:

$$1 = p(\varnothing|a) + \sum_x p(x|a) + \sum_{xy} p(xy|a).$$

However, we can only observe that it is one of the many possible solutions to this system. While dealing with generating functions, we are looking for the solution that is analytic at $z = 0$. Moreover, this analytic solution is unique and can be obtained by coefficient-wise iteration of the system of functional equations.

Could we just skip all these technical details and pretend we have what we need? In fact, no, and this detail will play an important role in identifying the transition between ergodicity and transience.

Now, let us look at the matrix $(\mathbf{I} - z\mathbf{P} - z\mathbf{K}(z))$ at $z = 1$ by substituting the above solution in the case when the alphabet has only two letters a and b:

$$\begin{pmatrix} 1 - p(a|a) - p(aa|a) - p(ab|a) & -p(b|a) - p(ba|a) - p(bb|a) & -p(\varnothing|a) \\ -p(a|b) - p(aa|b) - p(ab|b) & 1 - p(b|b) - p(ba|b) - p(bb|b) & -p(\varnothing|b) \\ -p(a|\varnothing) & -p(b|\varnothing) & 1 - p(\varnothing|\varnothing) \end{pmatrix}$$

It is clear that there is a linear dependence between the columns of this matrix: their sum gives a zero vector. We arrive to the following conclusion.

Proposition 2. If the analytic at zero solution $H_a(z)$ is the "trivial" branch $H_a(1) = 1$, then the Markov Chain is ergodic.

Proof. The inverse matrix turns to infinity, therefore, $G_{xy}(1) = \infty$. □

Suppose we are given the array of transition probabilities. How do we understand whether the Markov Chain is ergodic or not? Here comes to help the geometry of the solution space of the system of functional equations.

Crossing solutions: example. Consider a quadratic polynomial in x parametrised by its two roots:

$$f_{a,b}(x) = (x-a)(x-b).$$

The coalescence of the solutions happens when $a = b$. However, if we are only looking at its coefficients, and don't know the roots, there is still a way to identify a multiple solution:

$$f_{a,b}(z) = 0 \text{ and } \partial_x f_{a,b}(z) = 0 \quad \Leftrightarrow \quad a = b.$$

Imagine the situation in which a is constant and b is moved around. Also imagine that a is known and b is unknown.

The "analytic at zero" solution would correspond to identically $x = a$ until the point where the two solutions meet, when afterwards, it switches to $x = b$. We can reformulate the above condition as follows:

$$\boxed{f'_{a,b}(a) < 0 \quad \Leftrightarrow \quad a < b}$$

Indeed, we could manually check that $f'_{a,b}(a) = (2x - (a+b))|_{x=a} = a - b$.

Inspired by this simple example, we can do the same thing, but with matrices: the switching between the ergodic regime and transient happens when the Jacobian of the system of equations

$$H_a(z) = zp(\varnothing|a) + z\sum_x p(x|a)H_x(z) + z\sum_{xy} p(xy|a)H_y(z)H_x(z)$$

has zero determinant when $z = 1$ when we substitute the above trivial solution . This yields

$$\det\left(\delta_{xy} - p(y|x) - \sum_i p(iy|x) - \sum_j p(yj|x)\right)_{xy} = 0.$$

With some more additional work we could show that for ergodic walks this determinant is necessarily positive. However, unfortunately, positiveness of the determinant does not give us any guarantees.

Would it be possible that the above condition is reformulated in terms of dominant eigenvalues? Here is a simple conjecture.

Conjecture. The Markov Chain is ergodic if and only if

$$\lambda^*\left(p(y|x) + \sum_i p(iy|x) + \sum_j p(yj|x)\right)_{xy} < 1.$$

2.2. **Probabilistic view.** It turns out there is a very simple proof of the above conjecture without even introducing generating functions.

Let E_a denote the expected time to remove a letter a on the top of the word, which is defined formally as

$$E_a = \mathbb{E}\min\{t\colon |X_t| < n,\ |X_0| = n,\ X_0 = w_1w_2\dots w_{n-1}a\}.$$

In fact, E_a can be explicitly expressed as

$$E_a = H'_a(1).$$

This quantity satisfies a system of linear equations

$$\begin{aligned} E_a &= p(\varnothing|a) + \sum_x p(x|a)(1+E_x) + \sum_{xy} p(xy|a)(1+E_y+E_x) \\ &= 1 + \sum_x p(x|a)E_x + \sum_{xy} p(xy|a)(E_y+E_x). \end{aligned}$$

This can be written in the matrix-vector form:

$$\vec{E} = \vec{1} + \mathbf{A}\vec{E},$$

where $\mathbf{A}$ is a positive matrix. The matrix $\mathbf{A}$ has a unique positive eigenvalue λ. The system has a positive finite solution if and only if $\lambda < 1$.

3. The case of general d.

The case of general d can be reduced to $d = 2$. With a recoding trick. We introduce a mega-letter, which includes at most d usual letters. Then, modifying at most d last letters can be done with modification of at most two mega-letters.

Now we can follow the same path, but show more general equations for paths. Instead of the functions $H_a(z)$ we now need triparametric functions $H_{ab,c}(z)$ which enumerate walks with two top letters a, b and finishing with a length reduced by 1 with the top letter c. These functions satisfy a system of equations

$$H_{ab,c}(z) = zp(c|ab) + z\sum_{xy} p(xy|ab)H_{xy,c}(z) + \sum_{xyz} p(xyz|ab)\sum_g H_{ef,g}(z)H_{dg,c}(z).$$

Now, the only probabilistically based observation would be that if the walk is ergodic then for every a, b one has

$$\sum_c H_{ab,c}(1) = 1$$

because it should eventually diminish its height by 1. Apart from this observation, there is no guarantee for a simple explicit solution in this case.

The Green functions now satisfy a very similar system of equations:

$$G_{xy}(z) = \delta_{xy} + zp(\varnothing|x)G_{\varnothing y}(z) + z\sum_a p(a|x)G_{ay}(z) + z\sum_{ab}\sum_c p(ab|x)H_{ab,c}(z)G_{cy}(z),$$

which has again a compact matrix form

$$\mathbf{G}(z) = \mathbf{I} + z\mathbf{P}\mathbf{G}(z) + z\mathbf{K}(z)\mathbf{G}(z),$$

where

$$\mathbf{K}(z) = \left(\begin{cases} \sum_{ab} p(ab|x)H_{ab,y}(z), & \text{if } |y| = 1 \\ 0, & \text{if } |y| = 0 \end{cases}\right)_{xy}.$$

Ergodicity criterion. Idea. Define a matrix $\vec{p} = (p(\gamma, \delta))$, $|\gamma| = d$, $|\delta| < d$. This matrix defines the probability that starting from a word of length d we shall ever return to a word of a smaller length.

For recurrent chains, we have

$$\sum_\delta p(\gamma, \delta) = 1.$$

Then, we can write a system of non-linear equations defining the matrix p. This system takes form

$$\vec{p} = F(\vec{p}).$$

We need to identify a positive solution to the above equation (there might be several).

It can be proven, using fixed-point theorem, that there always exists a solution $\vec{p}_*$ satisfying for every γ

$$\sum_{\delta} p_*(\gamma, \delta) = 1.$$

Again, there might be several solutions satisfying this criterion. Without going into too much detail, say that there is a way to identify the "true" solution: it will be coordinatewise minimal.

We are one step from obtaining the ergodicity criterion using the dominant eigenvalue property. Define, as in the case $d = 1$,

$$E_\gamma = \mathbb{E}[\tau_{\gamma,n} \mid X_0 = \rho\gamma], \quad |\gamma| = d, |\rho\gamma| = n,$$

that is, the waiting time to remove one letter on condition that the top d letters are written as γ. These quantities satisfy a linear system of equations

$$\vec{E} = \vec{1} + \mathbf{A}(\vec{p}_*)\vec{E}.$$

Therefore, the process is ergodic if and only if the maximal eigenvalue λ of the matrix $\mathbf{A}(\vec{p}_*)$ is less than one.

4. Conclusion

What else can be studied using these tools apart from ergodicity?

- Stack behaviour: stack height. There is a theorem that
$$\lim_{n\to\infty} L_n/n = C > 0$$
- Sequence stabilization
- Other walk models. The space for generalisation is huge: queue where words can be modified from both sides, or if transition probability depend on something else (on the height of the stack, for example, or on the quantity of certain letters in the stack).
- Other generalisations: walks on graphs; non-deterministic versions of walk choices (existence probability, CSP).