SUMMER PROJECT'18

ON MIXING TIME OF THE JUGGLING CHAIN

Soham Bonnerjee

B.Stat Hons., Indian Statistical Institute, Kolkata

PROJECT GUIDE: Prof. Arvind Ayyer, Dept. of Mathematics, IISc Bangalore

July 10, 2018

Contents

INTRODUCTION

Juggling has been a fascinating human pastime since antiquity. We consider a mathematical modelling of Juggling procedure as established by *Warrington* [1], and further studied by *Prof. Arvind Ayyer, Jeremie Bouttier, Sylvie Corteel, Francois Nunzi* [2], giving it a Markov Chain model. The goal of this project has been to provide some insights on the Mixing Time of this markov chain model, and provide bounds on the same.

THE JUGGLING CHAIN AND STATIONARY DISTRIBUTION

Consider a *perfect* juggler *Adam* (i.e he would catch all the balls) juggling l balls. Assume that time is discretized, i.e at each second, a ball in the air comes down 1 metre (say). Also, Adam can't throw beyond a specified height, say h metre, and he can't gather more than one balls at a time, so that he has to throw in such a way as to at most one balls is in his hand in a particular time..A ball in the air can only come down in the next second, and a ball in Adam's hand can be thrown at any of the $h-(l-1) = h-l+1$ positions (since in this case) there would be $l-1$ balls in the air where no ball would come down in the next second. We assume that Adam chooses the height to throw the ball from available heights from an *uniform distribution* over $0,1,...,h-l$.
Now we consider the positions of the balls at each time point. From Adam's hand to the topmost height, we label each height as $1,2,..,h$. So there will be $\binom{h}{l}$ possible different configurations of positions of balls. A ball is at i^{th} height iff it reaches Adam's hand after i seconds.
Consider a h-length binary string. At a particular time, if a ball is in the i^{th} height, we write 1 at the i^{th} position of the string, and if there's no ball in that height, we write 0. With this model, we can express each configuration of balls in the above juggling procedure as a *binary string* of length h, consisting of l 1's, corresponding to l balls, and $h-l$ 0's. At each time point, any 1 that is not *leftmost*, shifts to the left. And if there is a leftmost 1, it goes any of the $h-l+1$ heights uniformly. Clearly this is a *Markov Chain* model.
To put mathematically, we follow Warrington's notations. let St_h denote the set of words of length h on the alphabet $1,0$, and let $St_{h,f} \subset St_h$ be the subset of words containing exactly $h-l = f$ occurrences of 0 . For $A \in St_{h,f}$ and $i \in 0,...,f$, we let $T_i(A) \in St_{h,f}$ be the word obtained by replacing the $(i+1)$-th occurrence of 0 in A by 1.

Definition 2.1. Given $h,l \in \mathbb{N}$, $h \geq l$, $h-l = f$, the ***Juggling Markov Chain*** is the Markov Chain on the state space $St_{h,f}$, with transition probability from $A = a_1a_2...a_h$ to B being:

$$P_{A,B} = \begin{cases} 1, & \text{if } a_1 = 0 \text{ and } B = a_2a_3...a_h0 \\ \frac{1}{f+1}, & \text{if } a_1 = 1 \text{ and } B = T_i(a_2a_3...a_h0) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

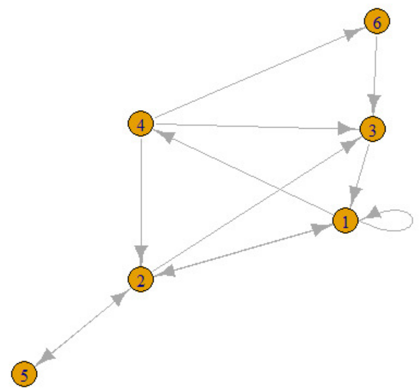

Figure 1: The Markov chain with $h = 4$ and $l = f = 2$

Figure 1. depicts the Markov Chain in the case $h = 4$, $l = 2$, and in the order of states $(1,1,0,0)$, $(1,0,1,0)$, $(0,1,1,0)$, $(1,0,0,1)$, $(0,1,0,1)$, $(0,0,1,1)$ (from 1 to 6), the transition matrix is:

$$P = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \tag{2}$$

Proposition 2.2 *This chain has a unique stationary distribution*

Proof: Consider the state $X = 1^l 0^f$. Clearly any state can reach this state by either inserting 1 at the lowest available 0 with probability $\frac{1}{f+1}$ or by getting to the only available state with probability 1 . Again, this state has a *self-loop*. Hence there exists a *unique closed irreducible* class having all states *aperiodic*.

Next, we prove that from X we can reach any state $A = a_1 a_2 ... a_h$. For each $i \in 0, 1, 2, .., h$, let n_i be the number of 0's in $a_1 a_2 ... a_i$, and define $A_i = 1^{l-i+n_i} 0^{f-n_i} a_1 ... a_i$, $A_0 = X$, $A_h = A$.Clearly $l \geq i - n_i$. Note that,

$$P_{A_i, A_{i+1}} = \begin{cases} 1, & \text{if } l = i - n_i \\ \frac{1}{f+1}, & \text{if } l > i - n_i \end{cases} \tag{3}$$

And thus we can reach any state from X within h steps, thus proving that that this chain is *irreducible* and aperiodic and hence by *Convergence Theorem* unique stationary distribution exists. ■

As Warrington showed *here*, the Stationary distribution, for a state v, is given by:

$$\pi(v) = \frac{\Delta(v)}{\left\{ {h+1 \atop f+1} \right\}} \tag{4}$$

$$\text{where } \Delta(v) = \prod_{t=1}^{h}(1+\phi_t(v)) \tag{5}$$

$$\phi_t(v) = \begin{cases} |t<j<h : v_j = 0|, & \text{if } v_t = 1 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Mixing Time

In a layman's terms, *Mixing Time* gives the rate of convergence of a Markov Chain to its Stationary Distribution, given it exists. It's formally defined using *Total Variation Distance* of the distribution of a chain starting from a particular state from the stationary distribution.

Definition 3.1: Let $d(t) := \max_{x\in St_{h,f}} ||P^t(x,\cdot) - \pi(\cdot)||_{\text{TV}}$. The ***Mixing Time*** of the Juggling Chain is given by:

$$t_{mix}(\epsilon) = \min\{t : d(t) < \epsilon\} \tag{7}$$

and

$$t_{mix} = t_{mix}(\frac{1}{4}) \tag{8}$$

Lower Bound on the Mixing Time of Juggling Chain

Proposition

$$t_{mix}(\epsilon) \geq \frac{\log(\binom{h}{l}(1-\epsilon))}{log(1+f)} \tag{9}$$

Proof:

Given a state x, $\Delta_{out}(x) = |y : P(x,y) > 0|$, i.e $\Delta_{out}(x)$ gives the maximum *outer-degree* of a state when the chain is represented as a (directed) graph. Similarly let Ψ_x^t denote the set of all the states accessible from state x in t steps, i.e $\Psi_x^t = y : P^t(x,y) > 0$.

Let $\Delta := \max_{x\in St_{h,f}} \Delta_{out}(x)$. Clearly, $\Delta_{out} = f+1$, as for any $A = a_1a_2..a_h$, if $a_1 = 0$ it can go to only one possible state with probability 1, but if $a_1 = 1$, it has $f+1$ possible states to go in the next step. Also, by definition of Δ, $|\Psi_x^t| < \Delta out^t = (f+1)^t$, since Δ^t maximizes the outer-degree at each step, thereby maximizing the number of states it can reach in each step,and hence maximizing Ψ_x^t. Thus proportion of time a chain starting from x

spends at Ψ_x^t in the long run $= \pi(\Psi_x^t) < \frac{(f+1)^t}{\binom{h}{l}}$ $\forall t$. Also, by definition of Ψ_x^t, $P^t(x, \Psi_x^t) = 1$
Now using definition of TV norm,
$||P^t(x,\cdot) - \pi(\cdot)||_{\text{TV}} = \max_{A \subset St_{h,f}} |P^t(x,A) - \pi(A)| \geq P^t(x, \Psi_x^t) - \pi(\Psi_x^t) \geq 1 - \dfrac{(f+1)^t}{\binom{h}{l}} > \epsilon$
$\iff (f+1)^t < (1-\epsilon)\binom{h}{l}$
which implies, from the definition of mixing time,

$t_{mix}(\epsilon) \geq \frac{\log(\binom{h}{l}(1-\epsilon))}{log(1+f)}$,

thus completing our proof. ■

Upper Bounds on Mixing Time for Special Cases

Case 1: $\underline{l = h-1}$
Let's look at $(h,l) = (3,2)$. We consider *Independent Coupling* of two juggling chains starting from two different states. Let $E_{x,y}(\tau)$ denote the *Coupling Time* of two chains starting from x and y, respectively.
Then in this case, a pair of chain starting at $(1,0,1) and (0,1,1)$ may couple at $(1,1,0)$ in $textbfone$ step with probability=$probability\ that(1,0,1) goes\ to(1,1,0) \times probability\ that\ (0,1,1) goes\ to(1,1,0) = \frac{1}{2}$, otherwise $(1,0,1)$ goes to $(0,1,1)$ with probability $\frac{1}{2}$ and $(0,1,1)$ goes to $(1,1,0)$ with probability 1, and we get another coupling starting from $(1,1,0)$ and $(0,1,1)$. Thus, similarly, in the order of the states $(1,1,0),(1,0,1),(0,1,1)$ we have the following set of equations:

$$\begin{aligned} E_{2,3}(\tau) &= \frac{1}{2} + \frac{1}{2}E_{1,3}(\tau) \\ E_{1,3}(\tau) &= \frac{1}{2} + \frac{1}{2}E_{1,2}(\tau) \\ E_{1,2}(\tau) &= \frac{1}{4}(1 + E_{1,2}(\tau) + E_{1,3}(\tau) + E_{2,3}(\tau)) \end{aligned}$$

Solving this, yields $E_{1,2}(\tau) = E_{2,3}(\tau) = E_{1,3}(\tau) = 1$
Now we use the theorem[3]: *Suppose that for each pair of states* $x, y \in \Omega$ *there is a coupling* (X_t, Y_t) *with* $X_0 = x$ *and* $Y_0 = y$. *For each such coupling, let* τ *be the first time the chains meet. Then* $d(t) \leq \max_{x,y \in \Omega} P_{x,y}(\tau > t)$.
Using *Markov's Inequality* we have,
$d(t) \leq \max_{x,y \in St_{h,f}} P_{x,y}(\tau > t) \leq \dfrac{\max_{x,y \in St_{h,f}} E_{x,y}(\tau)}{t} = \dfrac{1}{t}$.

So $d(t_{mix}(\epsilon) - 1) > \epsilon \iff t_{mix}(\epsilon) \leq \frac{1}{\epsilon} + 1$

If $(h,l)=(4,3)$, then the set of equations obtained likewise,in the order of $(1,1,1,0),(1,1,0,1),(1,0,1,1),(0,1,1,1)$ is:

$$\begin{aligned}
E_{3,4}(\tau) &= \frac{1}{2}+\frac{1}{2}E_{1,4}(\tau)\\
E_{1,4}(\tau) &= \frac{1}{2}+\frac{1}{2}E_{1,2}(\tau)\\
E_{2,4}(\tau) &= \frac{1}{2}+\frac{1}{2}E_{1,3}(\tau)\\
E_{1,2}(\tau) &= \frac{1}{4}(1+E_{1,2}(\tau)+E_{1,3}(\tau)+E_{2,3}(\tau))\\
E_{1,3}(\tau) &= \frac{1}{4}(1+E_{1,4}(\tau)+E_{1,2}(\tau)+E_{2,4}(\tau))\\
E_{2,3}(\tau) &= \frac{1}{4}(1+E_{1,3}(\tau)+E_{1,4}(\tau)+E_{3,4}(\tau))
\end{aligned}$$

Solving this, yields $E_{1,2}(\tau)=E_{1,3}(\tau)=E_{1,4}(\tau)=E_{2,3}(\tau)=E_{2,4}(\tau)=E_{3,4}(\tau)=1$ which in turn yields $t_{mix}(\epsilon)\le\frac{1}{\epsilon}+1$
We end this case with an obvious conjecture:

Conjecture 5.1. For the juggling chain, if $l=h-1$, $t_{mix}(\epsilon)\le\frac{1}{\epsilon}+1$

Case 2: $\underline{h=2l}$
Observations and Data:
For $l=2$, $h=4$, the $||P^t(x,\cdot)-\pi(\cdot)||_{TV}$ for each $x\in St_{4,2}$, for each $t\ge 4$, in the order of $(1,1,0,0),(1,0,1,0),(0,1,1,0),(1,0,0,1),(0,1,0,1),(0,0,1,1)$ is:

0.3066667 0.4266667 0.6400000 0.5600000 0.7600000 0.8400000 $(t=1)$

0.1644444 0.2888889 0.3066667 0.4000000 0.4266667 0.6400000 $(t=2)$

0.07259259 0.12148148 0.16444444 0.28000000 0.28888889 0.30666667 $(t=3)$

0.06320988 0.06567901 0.07259259 0.10024691 0.12148148 0.16444444 $(t=4)$

Clearly for the above data, $\forall\, t\le h$, $||P^t(x,\cdot)-\pi(\cdot)||_{TV}$ is maximum when $x=0^l1^l$ and $t_{mix}(\frac{1}{4})=t_{mix}=4=h$
Similar is the data for $l=3$, $h=6$, in the order of $(1,1,1,0,0,0),(1,1,0,1,0,0),(1,0,1,1,0,0),(0,1,1,1,0,0),(1,1,0,0,1,0),(1,0,1,0,1,0),(0,1,1,0,1,0),(1,0,0,1,1,0),(0,1,0,1,1,0),(0,0,1,1,1,0),(1,1,0,0,0,1),(1,0,1,0,0,1),(0,1,1,0,0,1),(1,0,0,1,0,1),(0,1,0,1,0,1),(0,0,1,1,0,1),(1,0,0,0,1,1),(0,1,0,0,1,1),(0,0,1,0,1,1),(0,0,0,1,1,1)$:

0.5428571, 0.6114286, 0.6628571, 0.8171429, 0.6914286, 0.7342857, 0.8628571, 0.7857143, 0.8971429, 0.9228571, 0.7828571, 0.8142857, 0.9085714, 0.8514286, 0.9314286, 0.9485714, 0.8942857, 0.9542857, 0.9657143, 0.9771429 $\cdots$ (t=1)

0.3271429, 0.3985714, 0.5428571, 0.5428571, 0.4750000, 0.6114286, 0.6114286, 0.6628571, 0.6628571, 0.8171429, 0.5378571, 0.6914286, 0.6914286, 0.7342857, 0.7342857, 0.8628571, 0.7857143, 0.7857143, 0.8971429, 0.9228571 $\cdots$ (t=2)

0.1941071, 0.2746429, 0.2837500, 0.3271429, 0.3430357, 0.3551786, 0.3985714, 0.5428571, 0.5428571, 0.5428571, 0.4201786, 0.4437500, 0.4750000, 0.6114286, 0.6114286, 0.6114286, 0.6628571, 0.6628571, 0.6628571, 0.8171429 $\cdots$ (t=3)

0.1111384, 0.1251339, 0.1537500, 0.1941071, 0.2662500, 0.2637946, 0.2746429, 0.2614732, 0.2837500, 0.3271429, 0.3348214, 0.3352232, 0.3430357, 0.3329018, 0.3551786, 0.3985714, 0.5428571, 0.5428571, 0.5428571, 0.5428571 $\cdots$ (t=4)

0.07400112, 0.06134487, 0.07790179, 0.11113839, 0.08571429, 0.10014509, 0.12513393, 0.12824777, 0.15375000, 0.19410714, 0.26321429, 0.25880580, 0.26625000, 0.25822545, 0.26379464, 0.27464286, 0.25835938, 0.26147321, 0.28375000, 0.32714286 $\cdots$(t=5)

0.05889648, 0.04539621, 0.05056920, 0.07400112, 0.05834403, 0.05363281, 0.06134487, 0.07670759, 0.07790179, 0.11113839, 0.08675781, 0.08069475, 0.08571429, 0.09907645, 0.10014509, 0.12513393, 0.12591239, 0.12824777, 0.15375000, 0.19410714 $\cdots$ (t=6)

Again $t_{mix} = 6$
So in both cases the largest tv norm of the distribution of the chain starting from a state from the stationary distribution is largest when the state is $0^l 1^l$, and the first time the corresponding TV norm is $< \frac{1}{4}$ is when $t = h = 2l$.
Computer simulations show that we get the same conclusion from $h = 8$, $l = 4$ case and $h = 10$, $l = 5$ case.Thus we have two more conjectures:

Conjecture 5.2. For the juggling chain, if $h = 2l \leftrightarrow f = l$, $\forall t \leq h$, $||P^t(x,\cdot) - \pi(\cdot)||_{TV}$ is maximum when $x = 0^l 1^l$

Conjecture 5.3. For the juggling chain, if $h = 2l \leftrightarrow f = l$, $t_{mix}(\frac{1}{4}) = t_{mix} = h = 2l$

An idea of the proof has been using induction on the Transition Matrix P, but I've been unable to make much progress in that direction.

References

[1] Gregory S.Warrington. *Juggling Probabilities*
https://arxiv.org/pdf/math/0302257v1.pdf

[2] Prof. Arvind Ayyer, Jeremie Bouttier, Sylvie Corteel, Francois Nunzi. *Multivariate Juggling Probabilities*
https://arxiv.org/pdf/1402.3752.pdf

[3] David A. Levin, Yuval Peres,Elizabeth L. Wilmer. *Markov Chain and Mixing Times,* COROLLARY 5.3, P-65.