



mixing times for tadpoles

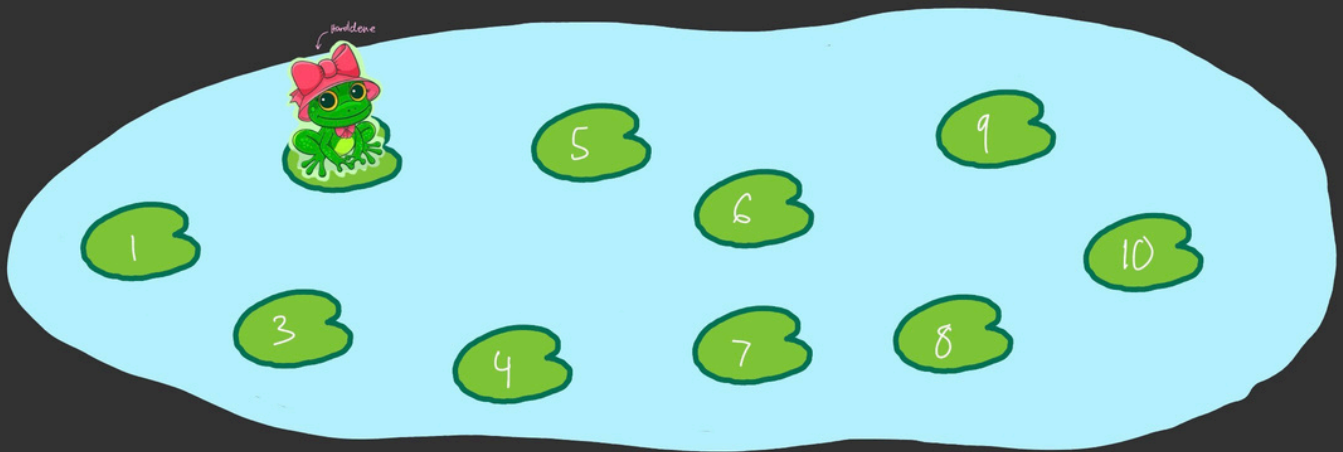


hopping to equilibrium

what the frog is this booklet?

if **haroldene the frog** hops around her pond for long enough, she'll know the exact **probability** of being at **each lily pad** in the **long run**.

mixing times explain how long it takes to get there.



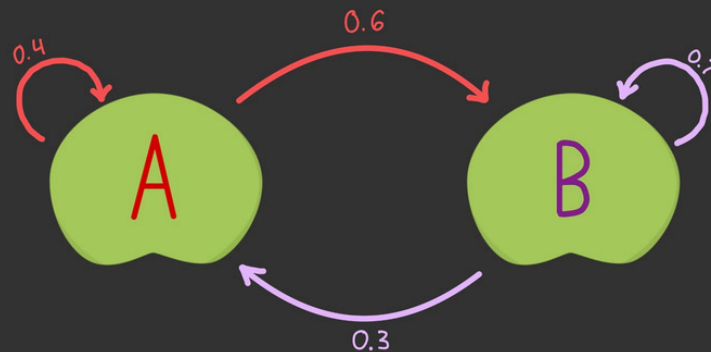
yea that's basically it.

there's a lot of ✨proofs and stuff✨ so hang tight.

skip past the markov chain and equilibrium intro stuff
if you been knew.

what the frog is a markov chain?

it's a bunch of **lily pads** (states) that haroldene can **hop** (transition) between. here we have two, but you haroldene can have as many as you she wants. the lily pads and hop probabilities can represent literally anything.



at each lily pad, haroldene can hop onto the same lily pad or another one. haroldene puts these **hop probabilities** into a **transition matrix** called **P**.

$$P = \begin{matrix} & \begin{matrix} \text{destination} \\ A & B \end{matrix} \\ \begin{matrix} \text{source} \\ A \\ B \end{matrix} & \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix} \end{matrix}$$

each **row** is for the lily pad haroldene is at **right now**. each **column** is for where she **will be** after one hop. so all the ways you can hop between them make up the entries of the matrix. for example, $P(A \rightarrow B) = 0.6$.

notice that the **#lily pads = #rows = #columns**.

also notice that the **numbers in each row add up to 1**, so each row is a **probability distribution** (i.e. list of ALL the possible things haroldene could do when she's at a particular lily pad and the corresponding probabilities)

what the frog is a markov chain?

out of her ♥♥cloaca♥♥, haroldene pulls out the chances she'll **start at A or B**.
we don't care how she pulled it out or why it looks like that, dw abt it.

you get the chances she's at **A or B after one hop** by **multiplying** that **OG probability distribution** by **P**. you multiply that new one by P again for the chances she's at A or B after two hops.

etc for however many hops her three-chambered heart desires.

$$\begin{array}{c} \text{after one hop} \\ \curvearrowright \\ \begin{bmatrix} a_0 & b_0 \end{bmatrix} P = \begin{bmatrix} a_1 & b_1 \end{bmatrix} \end{array}$$

haroldene pulls out a lil **example** out of her cloaca just for you. ♥

$$\begin{array}{c} \text{after one hop} \\ \curvearrowright \\ \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} 0.35 & 0.65 \end{bmatrix} \end{array}$$

notice that both the **OG** and **after-one-hop** distributions add up to 1.

(the chances haroldene is somewhere in the pond is always 100%.)

what the frog is equilibrium?

haroldene's on her **ironfrog grind**.

haroldene knows that after a long day of hopping, she'll know the **probabilities** of being at **A or B in the long run** (or long hop? what is the amphibian equivalent of a marathon? pls comment below thx).

this is the probability **distribution** where, **no matter how many hops** haroldene hops (#times she multiplies by P), the specific **probabilities of being at A and B won't change**.

$$\pi P = \pi$$

for you lovely freaky folx who took lin alg, it's the **eigenvector** w/ eigenvalue 1.

stays the same (after any number of hops)

$$\begin{bmatrix} a_\pi & b_\pi \end{bmatrix} \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} a_\pi & b_\pi \end{bmatrix}$$

if haroldene never gets stuck in any part of the pond and is never forced to go in cycles (depends on what P looks like), then she's chilling.

π will be unique! her one and only. 💍

what the frog is equilibrium?

how the frog does haroldene **find these numbers**? here are her freaky trix (literally just finding the eigenvector).

multiply the rhs by the identity matrix (changes nothing).

$$\begin{bmatrix} a_\pi & b_\pi \end{bmatrix} \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} a_\pi & b_\pi \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

get π on one side.

$$\begin{bmatrix} a_\pi & b_\pi \end{bmatrix} \left(\begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

now haroldene has a yummy new matrix.

$$\begin{bmatrix} a_\pi & b_\pi \end{bmatrix} \begin{bmatrix} -0.6 & 0.6 \\ 0.3 & -0.3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

she gets her glossal giddy on in the next page.

what the frog is equilibrium?

the equations haroldene get from π times the matrix are #twinning 🌹

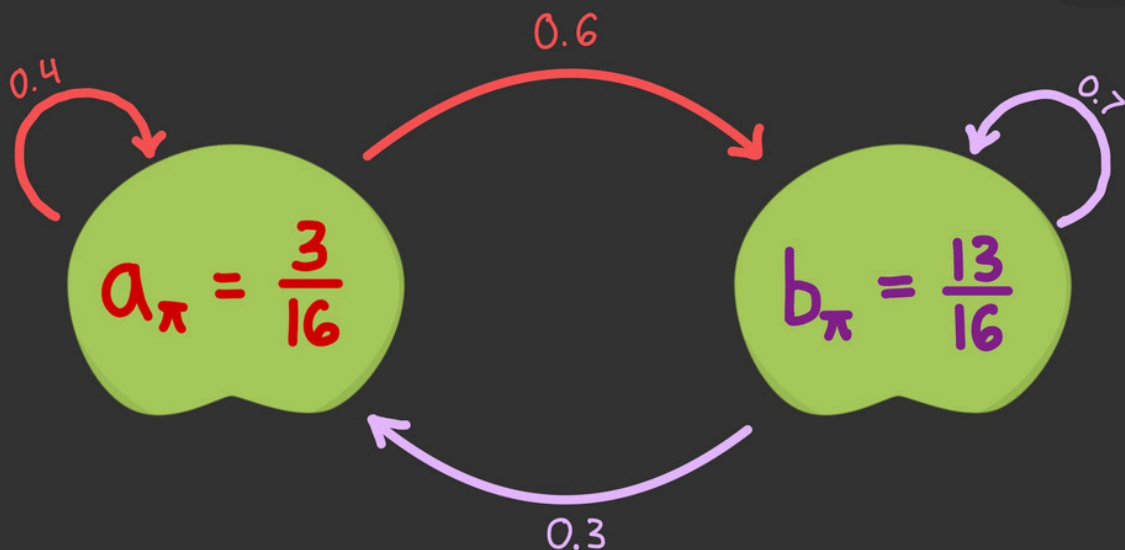
the fact that π is a probability distribution comes in clutch 🐱

$$-0.6 a_{\pi} + 0.3 b_{\pi} = 0$$

$$0.6 a_{\pi} - 0.3 b_{\pi} = 0$$

$$a_{\pi} + b_{\pi} = 1$$

haroldene does the algebra (shocker) and gets these values for the **probabilities** she'll be at **A** and **B** in the **long hop**. yippee!!

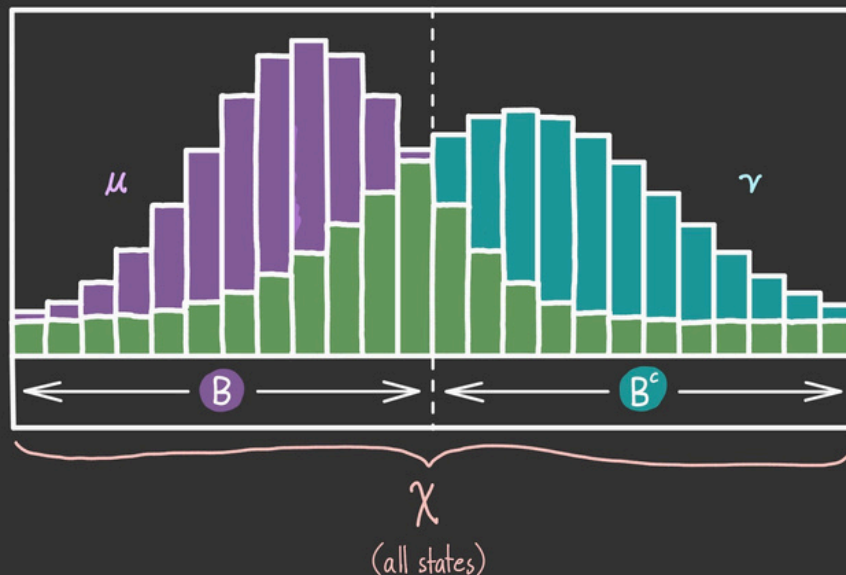


how far is she from equilibrium?

each row in P is a probability distribution. so is π .

we haroldene can graph any two probability distributions! she cares about how big the overlapping (green) and non-overlapping (purple/blue) areas are.

chi (fruity X) is the set of all lily pads. μ (μ) and ν (ν) are any two probability distributions.



this is our distribution μ (add up all the purple and green bars).

$$\text{purple square} + \text{green square} = 1$$

this is our distribution ν (add up all the blue and green bars).

$$\text{blue square} + \text{green square} = 1$$

how far is she from equilibrium?

purple and blue are both **1-green** (cut out the overlap from each curve).

so the **purple and blue areas** are the **exact same** 🐱



haroldene cares about how big this **non-overlapping region** because it'll tell her **how far** she is from her one true love (π). if purple/blue = 0, she's already united.

haroldene cares so much that she's talmboutsome "**total variation distance**". for which internalizing the **following eqns** is EXTREMELY IMPORTANT (#subtleforeshadowing).

$$\|\mu - \nu\|_{TV} = \sum_{x \in \mathcal{B}} [\mu(x) - \nu(x)] = \text{purple square}$$

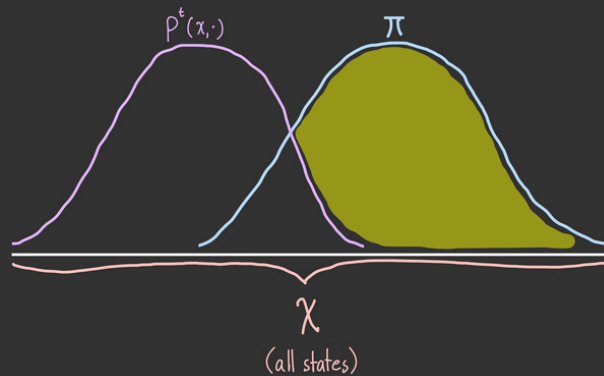
$$\|\mu - \nu\|_{TV} = \sum_{x \in \mathcal{B}^c} [\nu(x) - \mu(x)] = \text{teal square}$$

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|$$

$$\|\mu - \nu\|_{TV} = \text{purple square} = \text{teal square} = 1 - \text{green square}$$

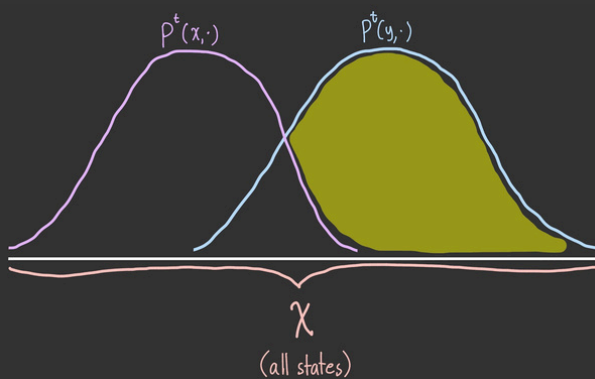
how far is she from equilibrium?

eventually each row in P^t (P after t hops) becomes π . haroldene **really cares** about the **row in P^t** that has the **biggest TVD** (smallest overlap) with π .



$$\max_{x \in \chi} \|P^t(x, \cdot) - \pi\|_{TV}$$

she also **really cares** about the **two rows in P^t** that have the **biggest TVD** between each other. such a caring soul or something like that ig.

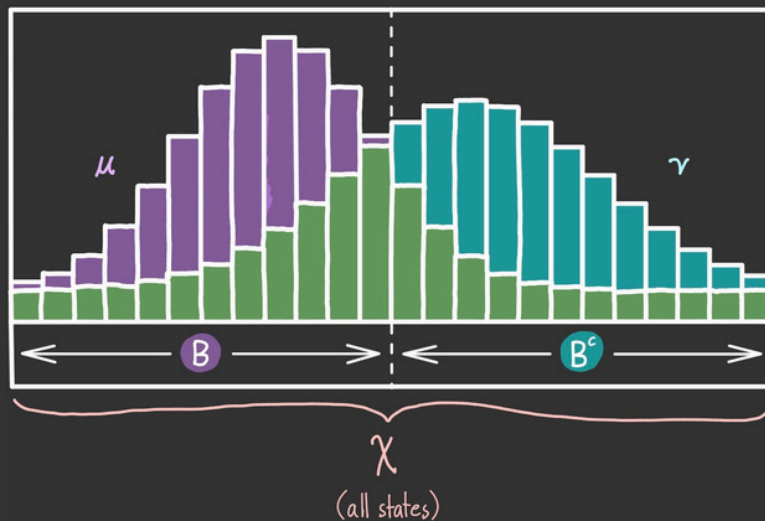


$$\max_{x, y \in \chi} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}$$

both of yellow **gaps** will **disappear** once **every row** = π (equilibrium reached).

haroldene in her #cupidera (?)

mommy frog (X) and daddy frog (Y) are a **couple**.



mommy frog (X)'s **probabilities** of being at each lily pad are laid out like **distribution μ** (purple + green).

the ENTIRE bar's **height** is the **probability**. the **x-coordinate** is the **lily pad**.

$$P\{X = x\} = \mu(x)$$

daddy frog (Y) is a ✨leech✨ so he copies mommy frog (X).

but he's #quirky so his probabilities are in **distribution ν** (blue + green).

$$P\{Y = y\} = \nu(y)$$

haroldene in her #cupidera (?)

haroldene is a (not so) recovering gambler.

she flips a coin with probability **green for heads** and **purple/blue for tails**.

$$\mathbb{P}\{X = Y\} = \text{green square}$$

$$\mathbb{P}\{X \neq Y\} = \text{purple square} = \text{teal square}$$

if haroldene gets heads, mommy frog and daddy frog **stay together** ($X=Y$).

if she gets tails, they get a **divorce** ($X \neq Y$).

makes sense bc they stay either together or don't (**green + purple/blue = 1**).

all this tomfroggerly was literally all just to get **this equation**:

$$\mathbb{P}\{X \neq Y\} = \|\mu - \nu\|_{TV}$$

haroldene will use this equation in part of her **#subtleforeshadowing** later 🐸

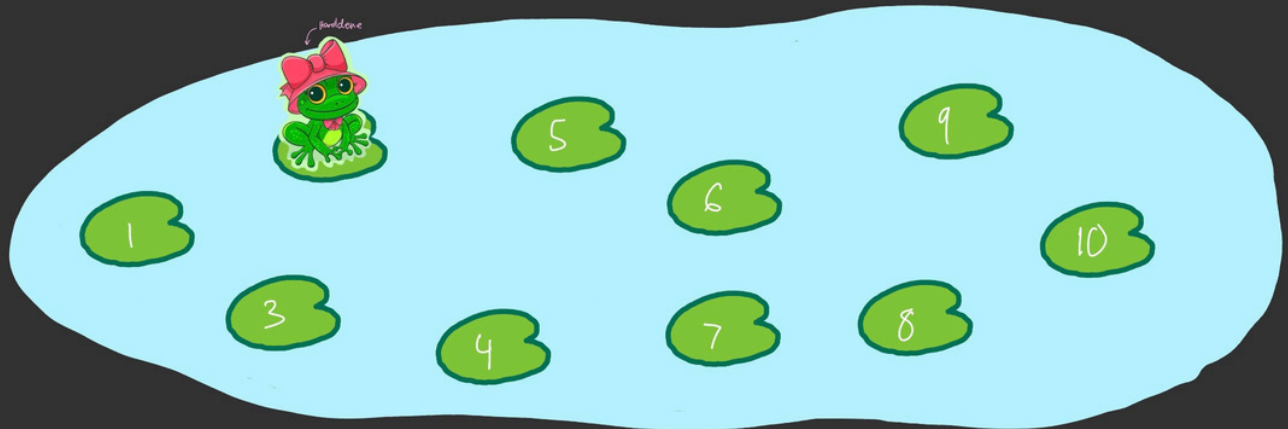
it will be in the "**submultiplicative property**" proof for haroldene's "**d-bar(t)**".

why the frog do we trust her?

haroldene is an **omniscient being** or something like that (idk the marsh magistrate didn't check her linkedin).

"it's only a matter of time..." haroldene sings miranda-esque-ly. until what?

haroldene says, give her enough hops, and she will **DEFINITELY** have a **fixed probability** of being at **each lily pad** in the **long run**.



perchance something **like this**:

$$\pi = \begin{bmatrix} 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.12 & 0.08 & 0.1 & 0.1 \end{bmatrix}$$

$\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4 \quad \pi_5 \quad \pi_6 \quad \pi_7 \quad \pi_8 \quad \pi_9 \quad \pi_{10}$

this is an **equilibrium** state. in the **long run**, she is certain that there's a **12% chance** she'll be at **lily pad 7** (ditto for the other lily pads and their probabilities).

"if you let me **hop around for long enough** (please), i'll **reach equilibrium for sure!**" she assures you. if you believe her already, she also has some **free candy** in her back of her **licenceless van** just for you, my friend. ❤️

why the frog do we trust her?

at **equilibrium**, the probability of going to a **particular lily pad** from **any other** is **constant**. $P(\text{any lily pad} \rightarrow \text{lily pad } i) = \pi_i$.

so each **column (col # = destination)** has the **same num** (π_i) all the way down.

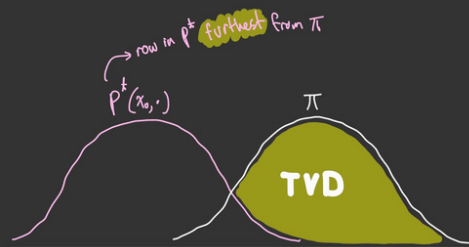
$$\Pi = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & \vdots & \cdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{bmatrix}$$

at **equilibrium**, P^t looks like Π . the **TVD (gap)** between **each row** and π is **0**. proving that the **BIGGEST gap (max TVD)** $\rightarrow 0$ guarantees this.

haroldene says she **can find a decreasing exponential** that the **max TVD** is always LEQ to (meaning $C > 0$ and α is between 0 and 1).

this makes **rhs** $\rightarrow 0$ as **t gets big** as frog (more hops). **lhs** $\leq$ **rhs** and can't be negative, so **lhs** $\rightarrow 0$ too.

$$\max_{x \in \mathcal{X}} \|P^t(x_0, \cdot) - \pi\|_{\text{TV}} \leq C \alpha^t$$



the **GOAL** is to show that **C** and α actually **EXIST**. in fact, they are hiding between her salmon sunhat and **slimy BALD head**. now hang tight...

why the frog do we trust her?

haroldene can get from **any lily pad to another** in **some number of hops (r)**.
(she never gets stuck in one part of the pond forever.)

so, after r hops, **everything in $P^r > 0$** . so, if δ is super **small**, this works:

$$P^r(x, y) \geq \delta \pi(y)$$

so δ **exists**. haroldene introduces her friend θ and says $\delta + \theta = 1$ (ok cool ig).

$$\delta + \theta = 1 \quad \delta = 1 - \theta \quad \theta = 1 - \delta$$

all this tomfroggerly fits together to make **this freaky eqn** (useful i swear).

$$P^r = (1 - \theta) \Pi + \theta Q$$

P^r and Π are **stochastic** (each **row adds up to 1**). haroldene shows **Q is too**.

$$P^r - (1 - \theta) \Pi = \theta Q$$

$$\frac{1}{\theta} P^r - \frac{(1 - \theta)}{\theta} \Pi = Q$$

$$(Q \text{ row sum}) = \frac{1}{\theta} (\overset{\text{stochastic}}{P^r \text{ row sum}}) - \frac{(1 - \theta)}{\theta} (\overset{\text{stochastic}}{\Pi \text{ row sum}})$$

$$Q \text{ row sum} = \frac{1}{\theta} (1) - \frac{1 - \theta}{\theta} (1)$$

$$Q \text{ row sum} = \cancel{\frac{1}{\theta}} - \cancel{\frac{1}{\theta}} + \frac{\theta}{\theta}$$

$$Q \text{ row sum} = 1$$

why the frog do we trust her?

everybody's so creative! or at least haroldene is bc now she's **putting a k in**.

$$P^{rk} = (1 - \theta^k) \top + \theta^k Q^k$$

she says that the **equation above works** for **any positive integer k**. to prove this, first she shows it's true for **k = 1**. this matches the **last pg's eqn** ✓

$$P^{r(1)} = (1 - \theta^1) \top + \theta^1 Q^1$$

$$P^r = (1 - \theta) \top + \theta Q$$

for haroldene's next trick, she'll show that if we **ASSUME** her equation **works** for **k = n**, it **must work** for **k = n + 1**.

so **assume** this is true:

$$P^{rn} = (1 - \theta^n) \top + \theta^n Q^n$$

and haroldene will **use the above** to **prove this**:

$$P^{r(n+1)} = (1 - \theta^{n+1}) \top + \theta^{n+1} Q^{n+1}$$

this is a **proof by induction** (not the stovetop, unless haroldene's peckish).

you can get onto the **first rung** (k=1) and climb onto each **next rung** (n → n+1).

why the frog do we trust her?

before showing we can climb our ladder, there are a **couple of things** haroldene wants you to know.

she has some **stochastic matrix** M (each row adds up to 1) and its **equilibrium distribution is** π .

$$\pi M = \pi$$

first thing: you can **replicate** the eqn above for **#(lily pads) rows** to get...

$$\Pi M = \Pi$$

second thing: **swapping** the order of M and Π **also gives** Π .

$$M \Pi = \Pi$$

to understand why, imagine the **dot product** between **M 's row** and **Π 's column** (needed for **matrix multiplication**). remember that the numbers in **M 's row add up to 1** and all the numbers in **Π 's column are the same**.

$$\begin{aligned} & [M \text{ row } i] \cdot [\Pi \text{ column } i] \\ &= [\text{nums summing to } 1] \cdot [\pi_i \dots \pi_i] \\ &= (\text{row sum} = 1) \pi_i \\ &= \pi_i \end{aligned}$$

why the frog do we trust her?

haroldene's on her webbed feet **begging you** to read this entire thing.

plsplspls it's not as deep as her pond, she's just **distributing** things and using the **eqns from the last two pages**. every little step is laid out.

$$\begin{aligned}
 P^{r(n+1)} &= P^{rn} P^r \\
 &= [(1-\theta^n)\Pi + \theta^n Q^n] P^r \\
 &= (1-\theta^n) \underbrace{\Pi \Pi \dots \Pi}_{\pi P = \pi, \text{ so } \Pi \Pi P = \Pi \text{ } r \text{ times}} P^r + \theta^n Q^n P^r \\
 &= (1-\theta^n) \Pi + \theta^n Q^n [(1-\theta)\Pi + \theta Q] \\
 &= (1-\theta^n) \Pi + \theta^n (1-\theta) \overbrace{Q^n \Pi}^{Q \Pi = \Pi \text{ } n \text{ times}} + \theta^n \theta Q^n Q \\
 &= (1-\theta^n) \Pi + (\theta^n - \theta^{n+1}) \Pi + \theta^{n+1} Q^{n+1} \\
 &= (1 - \cancel{\theta^n} + \cancel{\theta^n} - \theta^{n+1}) \Pi + \theta^{n+1} Q^{n+1} \\
 &= (1 - \theta^{n+1}) \Pi + \theta^{n+1} Q^{n+1} \\
 \therefore P^{r(n+1)} &= (1 - \theta^{n+1}) \Pi + \theta^{n+1} Q^{n+1}
 \end{aligned}$$

YAY!! this means we can **climb our ladder** ($n \rightarrow n+1$) and **get to the 1st rung**, so we can get to **any # rung** (so **k can be whatever** haroldene's three-chambered heart desires).

why the frog do we trust her?

okay cool so this is **true for any k** (big deal wtvr).

$$P^{rk} = (1 - \theta^k) \Pi + \theta^k Q^k$$

yes big deal acc bc plot twist(!) haroldene is **#sneaky** and will use this to get the **$P - \pi$ max TVD looking thing**.

she **multiplies both sides by P^j** and does some **#stealthy rearranging**.

$$P^{rk} P^j = (1 - \theta^k) \overbrace{\Pi \Pi}^{\Pi P = \Pi \text{ } j \text{ times}} P^j + \theta^k Q^k P^j$$

$$P^{rk+j} = (1 - \theta^k) \Pi + \theta^k Q^k P^j$$

$$P^{rk+j} = \Pi - \theta^k \Pi + \theta^k Q^k P^j$$

$$P^{rk+j} - \Pi = \theta^k Q^k P^j - \theta^k \Pi$$

$$P^{rk+j} - \Pi = \theta^k \underbrace{(Q^k P^j - \Pi)}_{\text{TVD at most 1}}$$

why the frog do we trust her?

from the last equation, take **any row** on each side and **the TVD**. $\text{TVD} \leq 1$ by defn so the **rhs is at most θ^k** .

the row is arbitrary so it **can be the max TVD row** (aka min overlap row).

$$\|P^{rk+j}(\overset{\text{any row (can be furthest)}}{\chi_0, \cdot}) - \pi\|_{TV} \leq \theta^k$$

θ is **positive** and **between 0 and 1** (exclusive) since $\delta + \theta = 1$.

so if haroldene **sets C and α** equal to the **following values** based on θ and r (which **MUST EXIST** since θ and r exist)...

$$C = \theta^{-1} \quad \alpha = \theta^{\frac{1}{r}}$$

...she knows $C > 0$ (reciprocal of +ve num) and $0 < \alpha < 1$ (root of 0-1 num).

plugging these values for **C and α** into haroldene's **exponential function**:

$$C \alpha^t = \theta^{-1} \theta^{\frac{t}{r}}$$

why the frog do we trust her?

since haroldene knows what she's replacing the rhs with, she can isolate k .

$$\Theta^k = \Theta^{-1} \Theta^{\frac{t}{r}}$$

$$k = \frac{t}{r} - 1$$

$$k = \frac{t-r}{r}$$

and plug it into the exponent of P so she has it in terms of t (num of hops).

$$rk + j = (t-r) + j$$

as t gets really big, haroldene hops more AND the rhs gets small as frog.
 every row in P must become π 🐸 #convergence #conformity
 #makethemarshmoistagain

so yuh haroldene trustworthy. she will reach equilibrium one hop(e)ful day 🌸

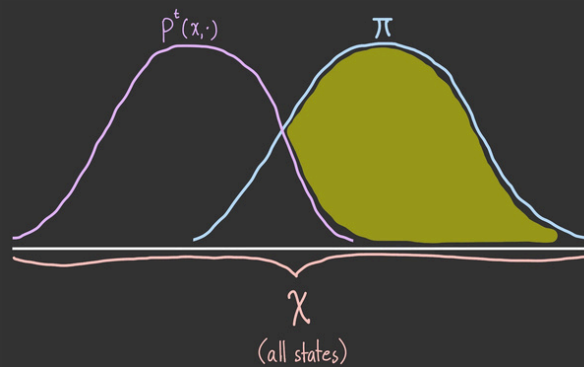
$$\underbrace{\| P^{(t-r)+j} (\chi_0, \cdot) - \pi \|_{TV}}_{\rightarrow 0} \leq \underbrace{\Theta^{-1} [\Theta^{\frac{1}{r}}]^t}_{\rightarrow 0} \quad \text{as } t \rightarrow \infty$$

haroldene's #subtleforeshadowing

haroldene is spooky with her ways and enjoys **foreshadowing** (un peu de **mise en place**, si vous voulez).

she calls **$d(t)$** the **biggest TVD** between **π** and **some row in P** after t hops.

$$d(t) := \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{TV}$$



she calls **$\bar{d}(t)$** the **biggest TVD** between **two rows in P** after t hops.

$$\bar{d}(t) := \max_{x, y \in \mathcal{X}} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}$$



once haroldene reaches **equilibrium**, every row in P will be π (no more yellow).

haroldene's #subtleforeshadowing

haroldene says the following **inequality** is super cool and **useful** (trust 🙏).

$$d(t) \leq \bar{d}(t) \leq 2d(t)$$



one must put their fly where their batrachian tongue is, so let's **prove it**.
starting with the **lhs**.

$$d(t) \leq \bar{d}(t)$$

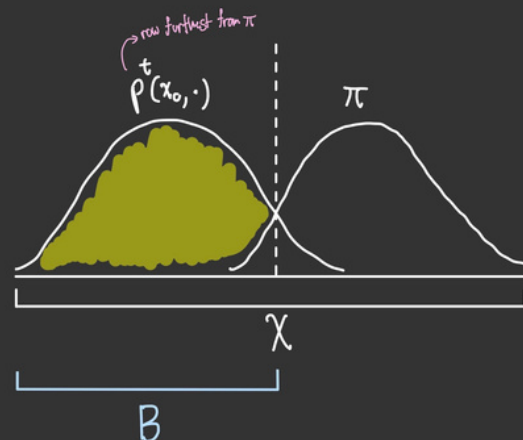
haroldene reminds you, talented tadpole, that each element in π is a **dot product** between π and a **column in P^t** .

she can get **all of π** this way or **pieces** of it (**A** can be 1, 2, or all the **lily pads**).

$$\pi = \pi P^t$$

$$\pi(A) = \sum_{y \in \chi} \pi(y) P^t(y, A)$$

$$A \subseteq \chi$$



let's say **row x_0** and **lily pads in B** give haroldene her **max distance from π** .

haroldene's #subtleforeshadowing

haroldene (sanders) is once again on her webbed feet asking you to **read through this proof** (every little step is laid out). the **yellow stuff** is from the **last page**. remember that all the **numbers in π** add up to 1.

$$\begin{aligned}
 d(t) &= \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{TV} \\
 &= |P^t(x_0, B) - \pi(B)| \quad \text{sum up differences over } B \\
 &= |P^t(x_0, B) - \sum_{y \in \mathcal{X}} \pi(y) P^t(y, B)| \\
 &= |1 P^t(x_0, B) - \sum_{y \in \mathcal{X}} \pi(y) P^t(y, B)| \\
 &= |\sum_{y \in \mathcal{X}} \pi(y) P^t(x_0, B) - \sum_{y \in \mathcal{X}} \pi(y) P^t(y, B)| \\
 &= |\sum_{y \in \mathcal{X}} [\pi(y) P^t(x_0, B) - \pi(y) P^t(y, B)]| \\
 &= |\sum_{y \in \mathcal{X}} \pi(y) [P^t(x_0, B) - P^t(y, B)]| \\
 &= \sum_{y \in \mathcal{X}} \pi(y) |P^t(x_0, B) - P^t(y, B)| \\
 &\leq \sum_{y \in \mathcal{X}} \pi(y) \|P^t(x_0, \cdot) - P^t(y, \cdot)\|_{TV} \\
 &\leq \sum_{y \in \mathcal{X}} \pi(y) \max_{x_0, y_0 \in \mathcal{X}} \|P^t(x_0, \cdot) - P^t(y_0, \cdot)\|_{TV} \\
 &= 1 \max_{x_0, y_0 \in \mathcal{X}} \|P^t(x_0, \cdot) - P^t(y_0, \cdot)\|_{TV} \\
 &= \bar{d}(t)
 \end{aligned}$$

haroldene's #subtleforeshadowing

now haroldene must **prove the rhs** of her beloved inequality chain.

$$\bar{d}(t) \leq 2 d(t)$$

she starts with the **defn of d-bar(t)**, adds a **0 in the middle**, uses the **triangle inequality** on each term, and links things back to **TVD** and **d(t)**.

$$\begin{aligned}\bar{d}(t) &= \frac{1}{2} \sum_{z \in \mathcal{X}} |P^t(x_o, z) - P^t(y_o, z)| \\&= \frac{1}{2} \sum_{z \in \mathcal{X}} |P^t(x_o, z) - \pi(z) + \pi(z) - P^t(y_o, z)| \\&= \frac{1}{2} \sum_{z \in \mathcal{X}} |(P^t(x_o, z) - \pi(z)) + (\pi(z) - P^t(y_o, z))| \\&\leq \frac{1}{2} \left(\sum_{z \in \mathcal{X}} |P^t(x_o, z) - \pi(z)| + \sum_{z \in \mathcal{X}} |\pi(z) - P^t(y_o, z)| \right) \\&= \frac{1}{2} \sum_{z \in \mathcal{X}} |P^t(x_o, z) - \pi(z)| + \frac{1}{2} \sum_{z \in \mathcal{X}} |P^t(y_o, z) - \pi(z)| \\&\leq \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\| + \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\| \\&= 2 \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\| \\&= 2 d(t)\end{aligned}$$

haroldene's #subtleforeshadowing

haroldene calls the inequality below $\bar{d}$'s "submultiplicative property" 🐱

$$\bar{d}(s+t) \leq \bar{d}(t) \bar{d}(s)$$

the **probability** of going from **lily pad x** to **lily pad w** is equal to the **expected value** of the **probability** that **t hops** can get haroldene to **w** .

$$P^{s+t}(x, w) = \sum_{z \in X} P(x_s = z) P^t(z, w) = \mathbb{E}(P^t(X_s, w))$$

haroldene's **freaky proof** below uses the fact above, the **$P(X \neq Y) = \text{TVD}$** thing from her #cupidera, and ofc the **defn of $\bar{d}$ -bar**. also, assume **x , y , and A** **maximize the TVD** between two rows in P^{s+t} .

$$\begin{aligned} \bar{d}(s+t) &= P^{s+t}(x, A) - P^{s+t}(y, A) \\ &= \mathbb{E}(P^t(x_s, A)) - \mathbb{E}(P^t(y_s, A)) \\ &= \mathbb{E}(P^t(x_s, A) - P^t(y_s, A)) \\ &\leq \mathbb{E}(\max_{x, y \in X} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \cdot 1_{\{X_s \neq Y_s\}}) \\ &= \max_{x, y \in X} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \mathbb{E}(1_{\{X_s \neq Y_s\}}) \\ &= \bar{d}(t) \|P^s(x, \cdot) - P^s(y, \cdot)\|_{\text{TV}} \\ &\leq \bar{d}(t) \max_{x, y \in X} \|P^s(x, \cdot) - P^s(y, \cdot)\|_{\text{TV}} \\ &= \bar{d}(t) \bar{d}(s) \end{aligned}$$

haroldene's #subtleforeshadowing

haroldene is a **#masterchef** in disguise. she uses her pantry to cook this up.

this is the **inequality chain** that will be directly useful getting upper bounds on mixing times (haroldene once again assures you, pls trust thx).

$$d(ct) \leq \bar{d}(ct) \leq \bar{d}(t)^c$$

the lhs is true bc of the whole $d(\text{🐼}) \leq d\text{-bar}(\text{🐼}) \leq 2d(\text{🐼})$ thing.

$$d(\text{😊}) \leq \bar{d}(\text{😊})$$

the rhs is true bc we just use the **submultiplicative** property **c times**.

$$\begin{aligned} \bar{d}(ct) &= \bar{d}(\overbrace{t + \dots + t}^{c \text{ of these}}) \\ &\leq \bar{d}(t)^1 \bar{d}(\overbrace{t + \dots + t}^{c-1 \text{ of these}}) \\ &\vdots \\ &\leq \bar{d}(t)^c \end{aligned}$$

what the frog is a mixing time?

took haroldene long enough to get here.

and by "here" i mean within an **error bound ε** of **convergence** to **equilibrium**.

$t_{\text{mix}}(\varepsilon)$ is the **number of hops** it took to **get that close** (TVD of ε or less).

$$t_{\text{mix}}(\varepsilon) := \min \{ t : d(t) \leq \varepsilon \}$$

haroldene defines t_{mix} as the **num hops** to get **within $\frac{1}{4}$** of **convergence**.

$$t_{\text{mix}} := t_{\text{mix}}\left(\frac{1}{4}\right) = \min \left\{ t : d(t) \leq \frac{1}{4} \right\}$$

her **cloaca** makes a comeback as this new delicious **inequality chain** comes out!

$$d(l t_{\text{mix}}(\varepsilon)) \leq \bar{d}(t_{\text{mix}}(\varepsilon))^l \leq (2\varepsilon)^l$$

l is a **positive integer** (so leftmost piece has a multiple of the mixing time).

the **lhs of this inequality chain** is true because of the inequality chain from the last page (holy russian nesting doll).

$$\begin{aligned} d(l t_{\text{mix}}(\varepsilon)) &\leq \bar{d}(l t_{\text{mix}}(\varepsilon)) \\ &\leq \bar{d}(t_{\text{mix}}(\varepsilon))^l \end{aligned}$$

what the frog is a mixing time?

the **rhs of this chain** is true bc of the rhs of $d(t) \leq \bar{d}(t) \leq 2d(t)$ and the fact that $d(t_{\text{mix}}(\epsilon)) \leq \epsilon$ by the **definition of $t_{\text{mix}}(\epsilon)$** .

$$\bar{d}(t_{\text{mix}}(\epsilon)) \leq 2d(t_{\text{mix}}(\epsilon))$$

$$\bar{d}(t_{\text{mix}}(\epsilon))^l \leq (2d(t_{\text{mix}}(\epsilon)))^l$$

$$\bar{d}(t_{\text{mix}}(\epsilon))^l \leq (2\epsilon)^l$$

using the **above** and the fact that she can make **l really big**, haroldene makes the **rhs = $(0-1 \text{ num})^l$** so that the **upper bound decreases** as **l increases**.

she chooses $\epsilon = \frac{1}{4}$ to make **rhs = $(0-1 \text{ num})^l$** .

$$\bar{d}(l t_{\text{mix}}(\epsilon))^l \leq (2\epsilon)^l$$

$$\epsilon = \frac{1}{4} \quad (t_{\text{mix}} := t_{\text{mix}}(\frac{1}{4}))$$

$$\bar{d}(l t_{\text{mix}})^l \leq (2 \cdot \frac{1}{4})^l$$

$$\bar{d}(l t_{\text{mix}})^l \leq (\frac{1}{2})^l$$

$$\bar{d}(l t_{\text{mix}})^l \leq 2^{-l}$$

what the frog is a mixing time?

haroldene's shiny slimy new **inequality chain** tells her:

$$d(\ell t_{\text{mix}}) \leq \bar{d}(\ell t_{\text{mix}})^{\ell} \leq 2^{-\ell}$$

so she has:

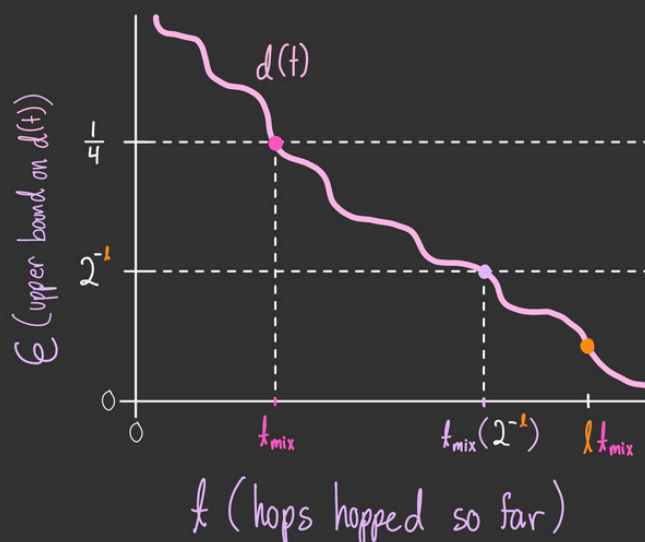
$$d(\ell t_{\text{mix}}) \leq 2^{-\ell}$$

which means the **TVD** after ℓt_{mix} hops is **smaller than** $2^{-\ell}$.

$$t_{\text{mix}}(2^{-\ell}) \leq \ell t_{\text{mix}}$$

this means the **# hops** it takes to get to a **TVD of** $2^{-\ell}$ is **less than** ℓt_{mix} .

haroldene is also a #masterartist and **explains this visually** below.



what the frog is a mixing time?

since **more hops** $\rightarrow$ **closer to equilibrium**:

$$t_{\text{mix}}(\epsilon) \leq t_{\text{mix}}(2^{-l}) \text{ as long as } 2^{-l} \leq \epsilon$$

if the rhs inequality is true, $t_{\text{mix}}(\epsilon)$ is also **smaller than** t_{mix} .

$$2^{-l} \leq \epsilon$$

$$-l \leq \log_2(\epsilon)$$

$$l \geq -\log_2(\epsilon)$$

$$[l \geq -\log_2(\epsilon)] \Rightarrow [t_{\text{mix}}(\epsilon) \leq t_{\text{mix}}(2^{-l}) \leq l t_{\text{mix}}]$$

first haroldene makes sure that l is in fact **bigger than** $-\log_2(\epsilon)$:

$$\text{Choose } l = \lceil -\log_2(\epsilon) \rceil$$

then she **plugs** l into the **rhs inequality** chain (ignore the middle piece):

$$t_{\text{mix}}(\epsilon) \leq \lceil -\log_2(\epsilon) \rceil t_{\text{mix}}$$

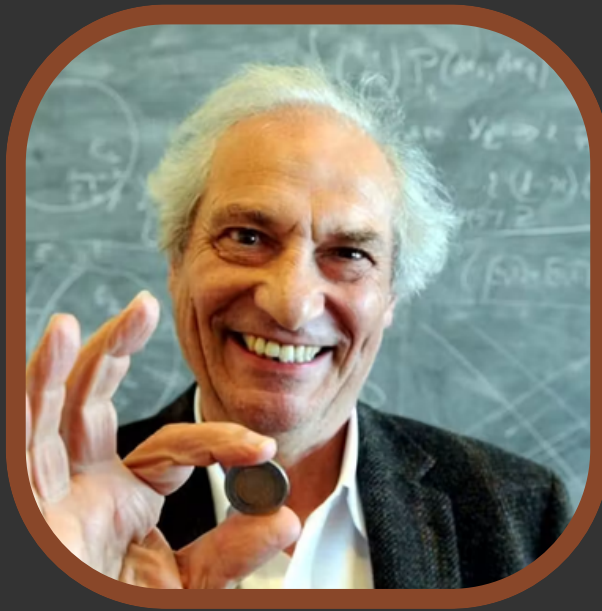
$$t_{\text{mix}}(\epsilon) \leq \lceil \log_2(\epsilon^{-1}) \rceil t_{\text{mix}}$$

haroldene the omniscient amphibian now knows **how many hops** it'll take **AT MOST** to get to her **desired closeness** (ϵ) to **equilibrium**.

so she can plan for her **ironfrog grind** accordingly...

persi diaconis (human + magician)

haroldene the frog has been leeching off of persi the human this entire time. in the late 1980s, persi diaconis and david aldous developed much of mixing time theory (and the marsh magistrate bought the rights).



one really cool application of mixing times (among tons, pls explore after this booklet) persi came up with is the fact that it takes 7 riffle shuffles to get a truly random arrangement of 52 playing cards.



a riffle shuffle is where you split a deck in half and perfectly interleave the two halves (alternating one card from each side).

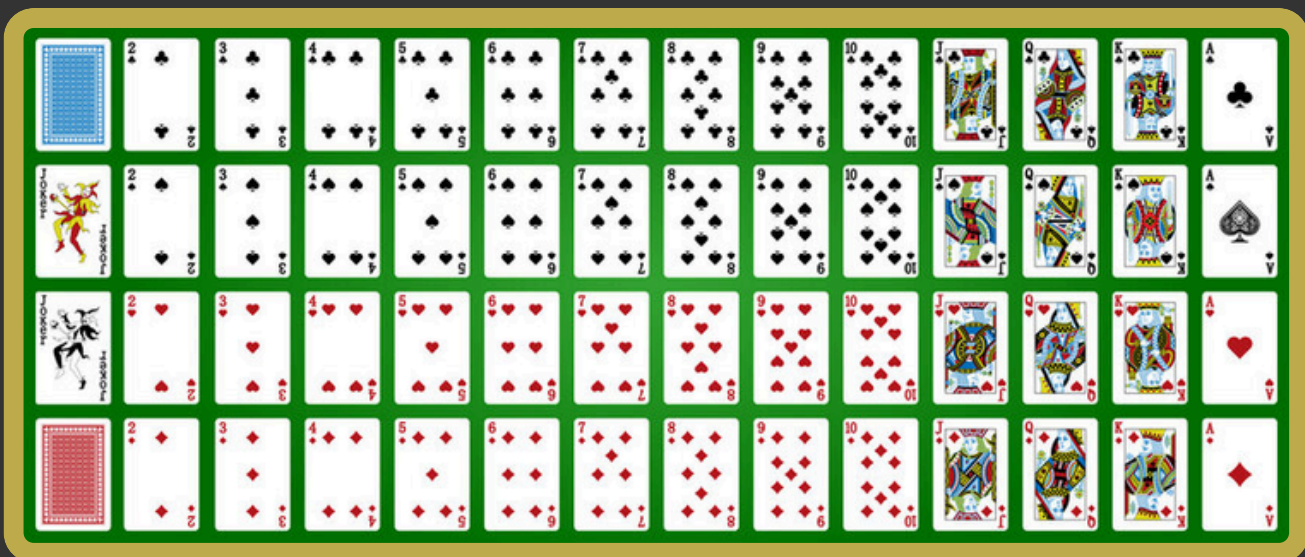
persi diaconis (human + magician)

haroldene knows what the deck looked like **starting off** (maybe a fresh deck).

between **1 and 4** shuffles, the ordering is still pretty **predictable**.

after **5** shuffles, it's getting **closer to random** but there's **still a pattern**.

by the **7th** shuffle, the **distance** from a truly **random arrangement** (TVD from equilibrium) **drops sharply**. this is called the "cutoff phenomenon".



rather than the distance from equilibrium shrinking gradually or linearly over time, the deck's **randomness remains severely compromised** until the **7th shuffle**, when the 52-card system tips into **complete disorder**.

further shuffles increase randomness marginally. **TVD** from equilibrium is **1** until the **7th shuffle** and then gets **extremely close to 0**.

why? haroldene is in her **retirement era** and makes you **explore on your own**.

persi diaconis (human + magician)

the gilbert-shannon-reeds (gsr) riffle shuffle model works by cutting a deck into two piles using a **binomial distribution** and **proportionally interleaving**.



persi diaconis and dave bayer used this model to show that, for an n -card deck, it takes about **$1.5 \cdot \log_2(n)$ shuffles** to get it to be truly random.

$$k_{\text{mix}} \approx \frac{3}{2} \log_2(n)$$

for a **52-card deck**, that gives haroldene:

$$\frac{3}{2} \log_2(52) \approx 8.25$$

this **translates to precisely 7 riffle shuffles** before the distance from equilibrium (i.e. truly random arrangements) gets super close to 0.

bye bye (or ribbit ig)

forgot everything haroldene told you?

"**equilibrium**" is when haroldene's **probability** of being at **each lily pad** stops changing (stays **fixed**).

that's when she has a **stationary distribution π** !

a **mixing time** is **how many hops** it takes haroldene to **get to equilibrium** (or a certain **distance from equilibrium**).

if haroldene knows **how close** she wants to be to **equilibrium**, she can find the **maximum number of hops** it takes to get there. #omniscient



haroldene likes putting her long tongue where her fly is, so she's **proved all the relevant equations** to you.

persi the magician was really the mastermind behind all this and also showed you some cool **applications to playing cards** (vegas calls)!

now that you've got a good foundation for the math behind markov chain mixing times, they both encourage you to **go and explore more applications!**



have a rad ribbity day!



special thanks to:
dr. noah forman

