On the Total Variation Distance of Labelled Markov Chains

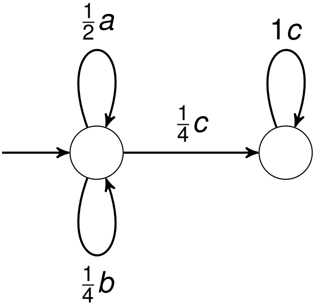
Taolue Chen[1] *Stefan Kiefer*[2]

[1]Middlesex University London, UK

[2]University of Oxford, UK

OASIS Seminar, Oxford
30 January 2015

Labelled Markov Chains (LMCs)

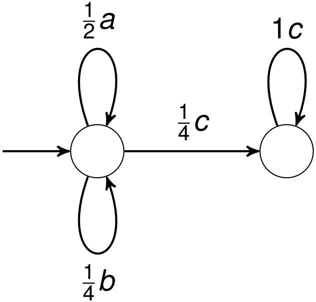


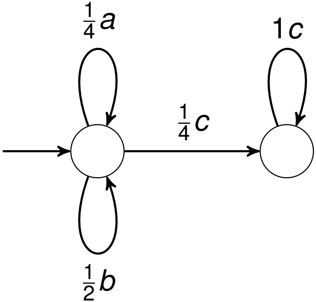
An LMC generates infinite words randomly.

$\Pr(\{abacccc\ldots\}) = \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{64}$
$\Pr(\{a\}\Sigma^\omega) = \frac{1}{2}$
$\Pr(\{b\}\Sigma^\omega) = \frac{1}{4}$
$\Pr(\text{“eventually only } c\text{”}) = 1$

Pr assigns a measurable event $E \subseteq \Sigma^\omega$ a probability $\in [0, 1]$.
$E \subseteq \Sigma^\omega$ could be defined by an LTL formula.

Labelled Markov Chains (LMCs)





$\Pr_1(\{accc\ldots\}) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$

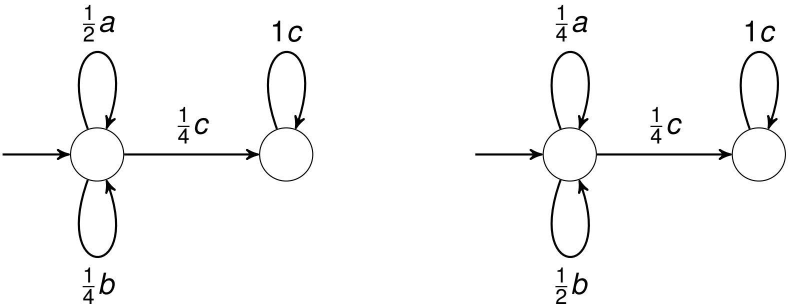
$\Pr_2(\{accc\ldots\}) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$

$\longrightarrow$ The two LMCs are not equivalent and have positive distance.

TV-distance $\boxed{d(\Pr_1, \Pr_2) := \max_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$

How large can $\Pr_1(E) - \Pr_2(E)$ get?

Labelled Markov Chains (LMCs)



$\Pr_1(\{accc\ldots\}) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$ $\qquad$ $\Pr_2(\{accc\ldots\}) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$

$\longrightarrow$ The two LMCs are not equivalent
and have positive distance.

TV-distance $\boxed{d(\Pr_1, \Pr_2) := \max_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$

How large can $\Pr_1(E) - \Pr_2(E)$ get? No larger than $\frac{15}{16} < 1$.

- "similar" LMCs $M_1, \ldots, M_n$
- (ω-regular) events $E_1, \ldots, E_m$
- want: bounds on $\Pr_i(E_j)$ for all i, j

Assume $d(\Pr_1, \Pr_i) \le \varepsilon$ for all $i \le n$. Then by definition

$$\forall i \; \forall j : \; \Pr_i(E_j) \in [\Pr_1(E_j) - \varepsilon, \Pr_1(E_j) + \varepsilon]$$

Digression: Total Variation Distance

			
$\Pr_{\text{Taolue}}$	0.3	0.6	0.1
$\Pr_{\text{Stefan}}$	0.2	0.5	0.3

$\Pr_{\text{Taolue}}\left(\left\{\right.$  $\left.\right\}\right) - \Pr_{\text{Stefan}}\left(\left\{\right.$  $\left.\right\}\right) = 0.1$

$\Pr_{\text{Taolue}}\left(\left\{\right.$  $\left.\right\}\right) - \Pr_{\text{Stefan}}\left(\left\{\right.$  $\left.\right\}\right) = 0.1$

$\Pr_{\text{Taolue}}\left(\left\{\right.$  ,  $\left.\right\}\right) - \Pr_{\text{Stefan}}\left(\left\{\right.$  ,  $\left.\right\}\right) = 0.2$

$\Pr_{\text{Taolue}}\left(\left\{\right.$  $\left.\right\}\right) - \Pr_{\text{Stefan}}\left(\left\{\right.$  $\left.\right\}\right) = -0.2$

Digression: Total Variation Distance

			
$\Pr_{\text{Taolue}}$	0.3	0.6	0.1
$\Pr_{\text{Stefan}}$	0.2	0.5	0.3

The TV-distance is half the L_1-norm of the difference:

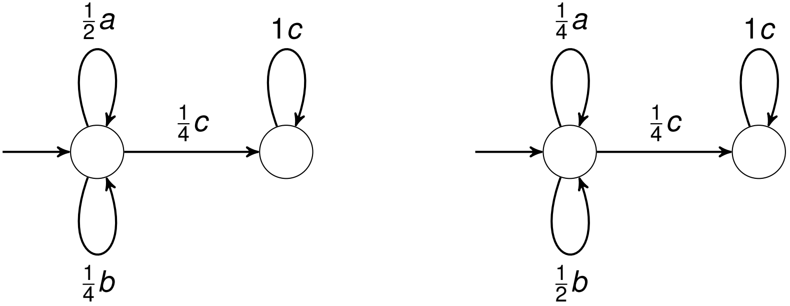
$$d(\Pr_{\text{Taolue}}, \Pr_{\text{Stefan}}) = \frac{1}{2} \|\Pr_{\text{Taolue}} - \Pr_{\text{Stefan}}\|_1$$

$$:= \frac{1}{2} \sum |\Pr_{\text{Taolue}}(x) - \Pr_{\text{Stefan}}(x)|$$

$x \in \{$ , ,  $\}$

$$= \frac{1}{2} \cdot (0.1 + 0.1 + 0.2) = 0.2$$

What is the Maximising Event?



$$\boxed{d(\Pr_1, \Pr_2) := \max_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$$

$d(\Pr_1, \Pr_2) = \Pr_1(E) - \Pr_2(E)$ holds for

$$E = \{wccc\ldots \mid w \in \{a,b\}^*,\ \#_a(w) \geq \#_b(w)\}$$

"*E* is a maximising event"

It's not clear that there is always a maximising event.

More Careful Definition

$$\boxed{d(\mathrm{Pr}_1, \mathrm{Pr}_2) := \max_{E \subseteq \Sigma^\omega} |\mathrm{Pr}_1(E) - \mathrm{Pr}_2(E)|}$$

More Careful Definition

$$\boxed{d(\Pr_1, \Pr_2) := \cancel{\max}_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$$

$$\boxed{d(\Pr_1, \Pr_2) := \sup_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$$

More Careful Definition

$$\boxed{d(\Pr_1, \Pr_2) := \cancel{\max}_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$$

$$\boxed{d(\Pr_1, \Pr_2) := \sup_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$$

Technically, E ranges only over the measurable subsets of Σ^ω (still uncountably many such events E).

More Careful Definition

$$\boxed{d(\Pr_1, \Pr_2) := \cancel{\max}_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$$

$$\boxed{d(\Pr_1, \Pr_2) := \sup_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$$

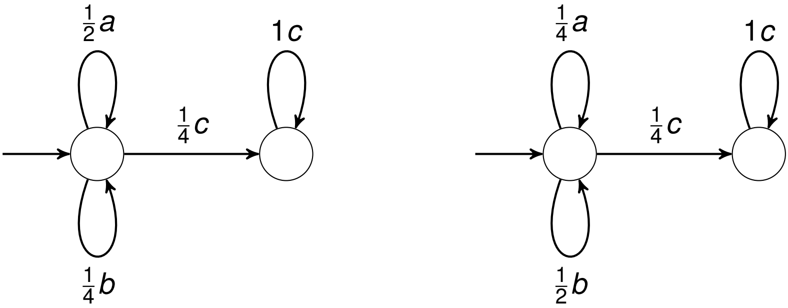
Technically, E ranges only over the measurable subsets of Σ^ω (still uncountably many such events E).

Proposition (Existence of a Maximising Event)

There is an event $E \subseteq \Sigma^\omega$ with $d(\Pr_1, \Pr_2) = \Pr_1(E) - \Pr_2(E)$.

$$\boxed{d(\Pr_1, \Pr_2) := \max_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$$

The Maximising Event is not Always ω-Regular.



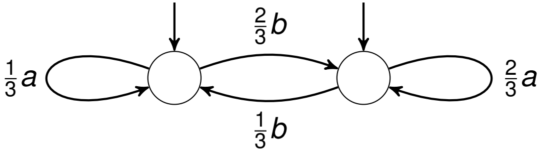
$$\boxed{d(\mathrm{Pr}_1, \mathrm{Pr}_2) := \max_{E \subseteq \Sigma^\omega} |\mathrm{Pr}_1(E) - \mathrm{Pr}_2(E)|}$$

$d(\mathrm{Pr}_1, \mathrm{Pr}_2) = \mathrm{Pr}_1(E) - \mathrm{Pr}_2(E) = \sqrt{2}/4$ holds for

$$E = \{wccc\ldots \mid w \in \{a,b\}^*,\ \#_a(w) \geq \#_b(w)\}$$

$\longrightarrow$ There is no ω-regular maximising event.

Example for Distance 1

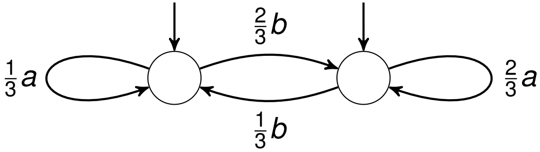


- The LMC is very symmetric.
- Both states enable all runs.

But $d(\Pr_1, \Pr_2) = 1$. What is the maximising event?

$b \quad a \quad a \quad b \quad b \quad a \quad a \quad b \quad b \quad a \quad b \quad a \quad b \quad a \quad b \quad b \quad \cdots$

Example for Distance 1

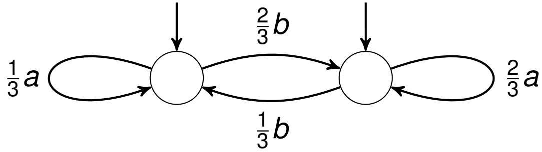


- The LMC is very symmetric.
- Both states enable all runs.

But $d(\Pr_1, \Pr_2) = 1$. What is the maximising event?

b	a	a	b	b	a	a	b	b	a	b	a	b	a	b	b	$\cdots$
1			2	3			4	5		6		7		8	9	

Example for Distance 1

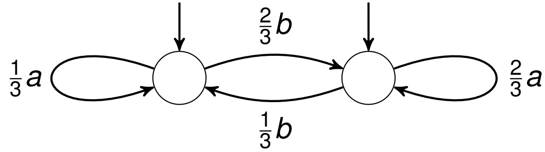


- The LMC is very symmetric.
- Both states enable all runs.

But $d(\Pr_1, \Pr_2) = 1$. What is the maximising event?

b	a	a	b	b	a	a	b	b	a	b	a	b	a	b	b	$\cdots$
1			2	3			4	5		6		7		8	9	
				b				b			a				b	
				$\frac{0}{1}$				$\frac{0}{2}$			$\frac{1}{3}$				$\frac{1}{4}$	$\to \frac{1}{3}$

Example for Distance 1



- The LMC is very symmetric.
- Both states enable all runs.

But $d(\mathrm{Pr}_1, \mathrm{Pr}_2) = 1$. What is the maximising event?

b	a	a	b	b	a	a	b	b	a	b	a	b	a	b	b	$\cdots$
1			2	3			4	5		6		7		8	9	
				b				b			a				b	
				$\frac{0}{1}$				$\frac{0}{2}$			$\frac{1}{3}$				$\frac{1}{4}$	$\to \frac{1}{3}$

Let $E =$ "this sequence converges to $\frac{1}{3}$". Then:

$$\mathrm{Pr}_1(E) = 1 \text{ and } \mathrm{Pr}_2(E) = 0$$

There is no ω-regular maximising event.

A Maximising Event: Intuition

			
$\Pr_{\text{Taolue}}$	0.3	0.6	0.1
$\Pr_{\text{Stefan}}$	0.2	0.5	0.3

$\Pr_{\text{Taolue}}\left(\left\{\right.\right.$ ,  $\left.\left.\right\}\right) - \Pr_{\text{Stefan}}\left(\left\{\right.\right.$ ,  $\left.\left.\right\}\right) = 0.2$

For LMCs, define

$$\overline{L}(w) := \frac{\Pr_2(w)}{\Pr_1(w)}$$

Maybe $E = \{w \in \Sigma^\omega \mid \overline{L}(w) \leq 1\}$ is a maximising event?

A Maximising Event: Intuition

			
$\mathrm{Pr}_{\mathrm{Taolue}}$	0.3	0.6	0.1
$\mathrm{Pr}_{\mathrm{Stefan}}$	0.2	0.5	0.3

$\mathrm{Pr}_{\mathrm{Taolue}}\left(\{$ ,  $\}\right) - \mathrm{Pr}_{\mathrm{Stefan}}\left(\{$ ,  $\}\right) = 0.2$

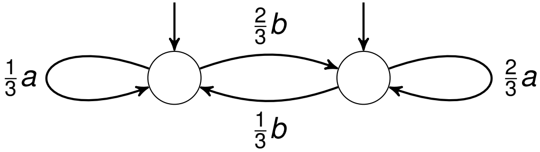
For LMCs, define

$$\overline{L}(w) := \frac{\mathrm{Pr}_2(w)}{\mathrm{Pr}_1(w)}$$

Maybe $E = \{w \in \Sigma^\omega \mid \overline{L}(w) \leq 1\}$ is a maximising event?

Redefine $\overline{L}(w) \ldots$

A Maximising Event



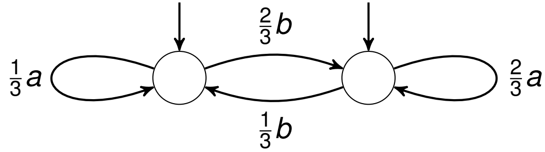
Fix a run $w = a_1 a_2 a_3 \cdots \in \Sigma^\omega$.
For every $k \in \mathbb{N}$ define a nonnegative value:

$$L_k(w) := \frac{\Pr_2(a_1 \cdots a_k \Sigma^\omega)}{\Pr_1(a_1 \cdots a_k \Sigma^\omega)}$$

w	b	a	a	b	$\cdots$
$\Pr_2(a_1 \cdots a_k \Sigma^\omega)$	$\frac{1}{3}$	$\frac{1}{3} \cdot \frac{1}{3}$	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3}$	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\cdots$
$\Pr_1(a_1 \cdots a_k \Sigma^\omega)$	$\frac{2}{3}$	$\frac{2}{3} \cdot \frac{2}{3}$	$\frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3}$	$\frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3}$	$\cdots$
$L_k(w)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\cdots$

If the run w is produced randomly (say, from the left state), $L_1, L_2, \ldots$ is a sequence of random variables.

A Maximising Event



Fix a run $w = a_1 a_2 a_3 \cdots \in \Sigma^\omega$.
For every $k \in \mathbb{N}$ define a nonnegative value:

$$L_k(w) := \frac{\Pr_2(a_1 \cdots a_k \Sigma^\omega)}{\Pr_1(a_1 \cdots a_k \Sigma^\omega)}$$

w	b	a	a	b	$\cdots$
$\Pr_2(a_1 \cdots a_k \Sigma^\omega)$	$\frac{1}{3}$	$\frac{1}{3} \cdot \frac{1}{3}$	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3}$	$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$	$\cdots$
$\Pr_1(a_1 \cdots a_k \Sigma^\omega)$	$\frac{2}{3}$	$\frac{2}{3} \cdot \frac{2}{3}$	$\frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3}$	$\frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3}$	$\cdots$
$L_k(w)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\cdots$

If the run w is produced randomly (say, from the left state), $L_1, L_2, \ldots$ is a sequence of random variables.

A Maximising Event

If w is produced randomly,
  $L_1, L_2, \ldots$ is a sequence of random variables.

For any prefix $a_1 \cdots a_k$:

$$\mathbb{E}_1 \left(L_{k+1}(w) \mid w \in a_1 \cdots a_k \Sigma^\omega \right) = L_k(w)$$

"$L_1, L_2, \ldots$ is a martingale"

A Maximising Event

If w is produced randomly,
$L_1, L_2, \ldots$ is a sequence of random variables.

For any prefix $a_1 \cdots a_k$:

$$\mathbb{E}_1 \left(L_{k+1}(w) \mid w \in a_1 \cdots a_k \Sigma^\omega \right) = L_k(w)$$

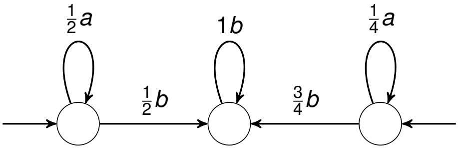
"$L_1, L_2, \ldots$ is a martingale"

The martingale is nonnegative.
$\Longrightarrow$ Martingale Convergence Theorem applies.
$\Longrightarrow$ $\overline{L} := \lim_{k \to \infty} L_k$ exists almost surely.
($\overline{L}$ is a random variable.)

Theorem (A Generic Maximising Event)

Define $E := \{w \in \Sigma^\omega \mid \overline{L}(w) \leq 1\}$.
Then $d(\Pr_1, \Pr_2) = \Pr_1(E) - \Pr_2(E)$.

Approximation: Lower Bound



$$\boxed{d(\Pr_1, \Pr_2) := \max_{E \subseteq \Sigma^\omega} |\Pr_1(E) - \Pr_2(E)|}$$

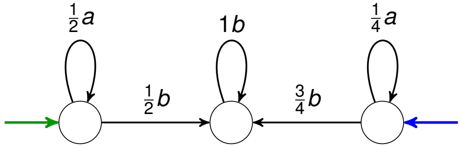
Fix $k \in \mathbb{N}$.
Idea: consider only events definable by the length-k prefix.
I.e., define $d_k(\Pr_1, \Pr_2) := \max_{W \subseteq \Sigma^k} |\Pr_1(W\Sigma^\omega) - \Pr_2(W\Sigma^\omega)|$.

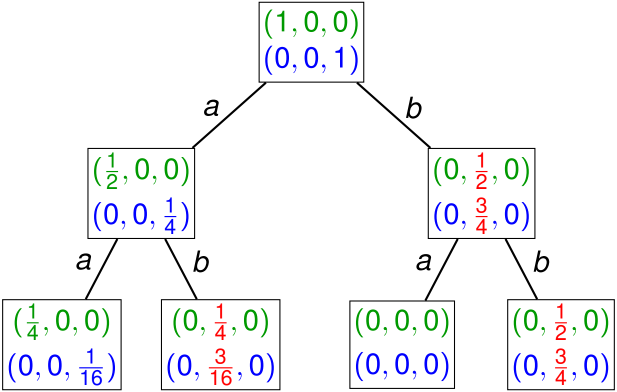
Proposition

For all $k \in \mathbb{N}$:

$$d_k(\Pr_1, \Pr_2) \leq d_{k+1}(\Pr_1, \Pr_2) \leq d_\infty(\Pr_1, \Pr_2) = d(\Pr_1, \Pr_2)$$

Approximation: Upper Bound





$0 =: \mathit{con}(0)$

$\frac{1}{2} =: \mathit{con}(1)$

$\frac{3}{16} + \frac{1}{2} =: \mathit{con}(2)$

Approximation: Upper Bound

This defines an increasing sequence:

$$0 \le con(0) \le con(1) \le \ldots \le con(\infty) = 1 - d(\mathrm{Pr}_1, \mathrm{Pr}_2)$$

In general, define $con(k)$ using equivalent distributions rather than equal states.

Approximation: Upper Bound

This defines an increasing sequence:

$$0 \le \mathit{con}(0) \le \mathit{con}(1) \le \ldots \le \mathit{con}(\infty) = 1 - d(\mathrm{Pr}_1, \mathrm{Pr}_2)$$

In general, define $\mathit{con}(k)$ using equivalent distributions rather than equal states.

Theorem

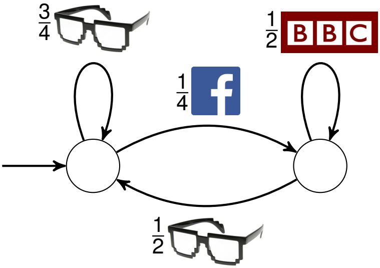
Given $\varepsilon > 0$, one can compute $x \in \mathbb{Q}$ with

$$d(\mathrm{Pr}_1, \mathrm{Pr}_2) \in [x, x + \varepsilon].$$

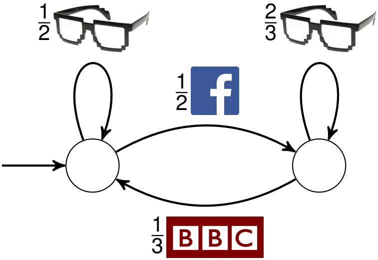
Open: convergence speed

The Distance-1 Problem: Deniability

Taolue:



Stefan:

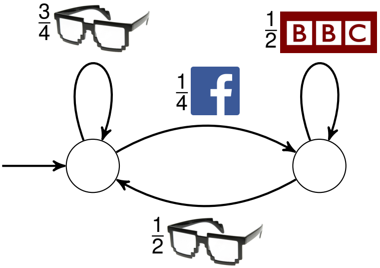


Task for NSA: distinguish those guys!

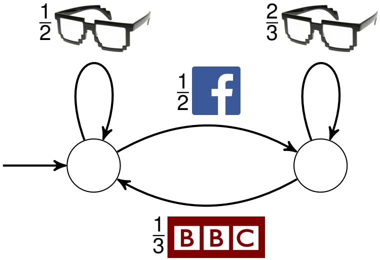
Is there $E \subseteq \{$  $\}^{\omega}$ with

$$\Pr_{\text{Taolue}}(E) = 1 \text{ and } \Pr_{\text{Stefan}}(E) = 0 \text{ ?}$$

Taolue:



Stefan:



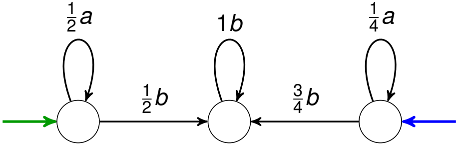
Task for NSA: distinguish those guys!

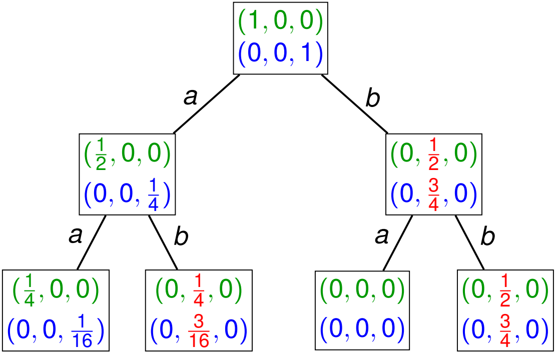
Is there $E \subseteq \{\ \ , \mathrm{f}, \mathrm{BBC}\}^\omega$ with

$$\Pr_{\text{Taolue}}(E) = 1 \text{ and } \Pr_{\text{Stefan}}(E) = 0 \text{ ?}$$

If not, Taolue can plausibly deny that he is Taolue.

The Distance-1 Problem





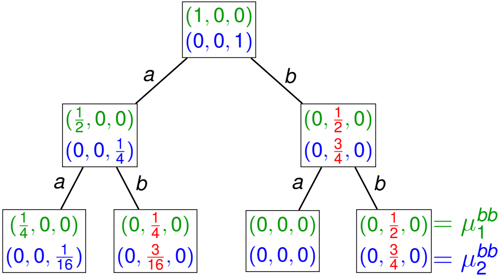
$0 =: \mathit{con}(0)$

$\frac{1}{2} =: \mathit{con}(1)$

$\frac{3}{16} + \frac{1}{2} =: \mathit{con}(2)$

The distance is $< 1 \quad \iff \quad \exists k \in \mathbb{N} : \mathit{con}(k) > 0$

The Distance-1 Problem: in PSPACE



The distance is $< 1 \quad \iff \quad \exists u \in \Sigma^* : \mu_1^u$ and μ_2^u overlap.

- Whether μ_1^u, μ_2^u overlap depends only on the supports of μ_1^u and μ_2^u.
- Whether μ_1^u, μ_2^u overlap can be computed in poly time using previous work and linear programming.
- There are at most $2^{2|Q|}$ possible supports of μ_1^u and μ_2^u.

$\longrightarrow$ PSPACE algorithm: guess a word $u \in \Sigma^{\leq 2^{2|Q|}}$ and check if μ_1^u, μ_2^u overlap.

The Distance-1 Problem: in P

To get a polynomial-time algorithm:

1. Generalise distance between states
   to distance between state distributions.
2. Exploit structural properties of the generalised notion:

Lemma

$$d(\pi_1, \pi_2) = 0 \quad \Longrightarrow \quad \forall q \in supp(\pi_1) : d(q, \pi_2) < 1$$
$$d(\pi_1, \pi_2) < 1 \quad \Longrightarrow \quad \exists q \in supp(\pi_1) : d(q, \pi_2) < 1$$

The Distance-1 Problem: in P

To get a polynomial-time algorithm:

1. Generalise distance between states
   to distance between state distributions.
2. Exploit structural properties of the generalised notion:

Lemma

$$
\begin{aligned}
d(\pi_1, \pi_2) = 0 \quad &\Longrightarrow \quad \forall q \in supp(\pi_1) : d(q, \pi_2) < 1 \\
d(\pi_1, \pi_2) < 1 \quad &\Longrightarrow \quad \exists q \in supp(\pi_1) : d(q, \pi_2) < 1
\end{aligned}
$$

3. Exploit previous work on LMC equivalence
   and use linear programming.

Theorem (Distance-1 Problem)

One can decide in polynomial time whether the distance between two LMCs is 1.

Threshold Problem
Input: 2 LMCs and threshold $\tau \in [0, 1]$
Output: Is $d(\Pr_1, \Pr_2) \geq \tau$?

Square-Root-Sum Problem
Input: $s_1, \ldots, s_n \in \mathbb{N}$ and $t \in \mathbb{N}$
Output: Is $\sum_{i=1}^{n} \sqrt{s_i} \geq t$?

The Square-Root-Sum problem is not known to be in NP.

Threshold Problem

Threshold Problem
Input: 2 LMCs and threshold $\tau \in [0, 1]$
Output: Is $d(\mathrm{Pr}_1, \mathrm{Pr}_2) \geq \tau$?

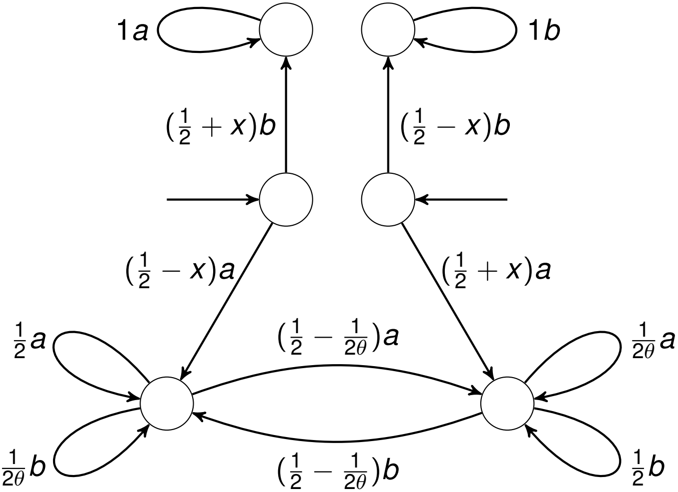
Square-Root-Sum Problem
Input: $s_1, \ldots, s_n \in \mathbb{N}$ and $t \in \mathbb{N}$
Output: Is $\sum_{i=1}^{n} \sqrt{s_i} \geq t$?

The Square-Root-Sum problem is not known to be in NP.

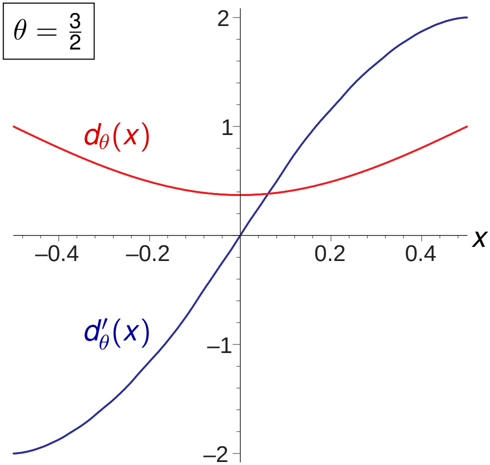
Theorem

The Threshold Problem is NP-hard.
The Threshold Problem is
hard for the Square-Root-Sum Problem.

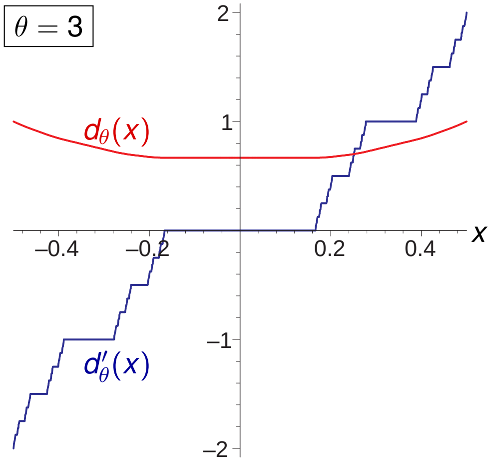
LMC with 2 Parameters



Distance as Function in x



Distance as Function in x



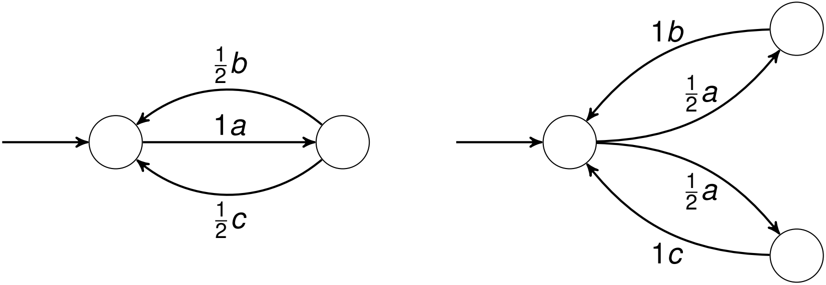
Bernoulli-Convolutions

- $d'_\theta(x)$ (rescaled) is the cumulative distribution function of

$$\sum_{i=0}^{\infty} \frac{X_i}{\theta^i} \qquad \text{with } \Pr(X_i = -1) = \Pr(X_i = +1) = \frac{1}{2}$$

- "Bernoulli convolutions": studied since the 1930s
- $\forall \theta > 1$: d'_θ is either absolutely continuous or singular.
- d'_3 is the (ternary) Cantor function.
- For almost all $\theta \in (1, 2]$: d'_θ is absolutely continuous.
- If θ is a Pisot number, then d'_θ is singular. [Erdős, 1939]
- It is open, e.g., whether $d'_{3/2}$ is absolutely continuous.

Related Work: Bisimilarity Pseudometric



LMCs are (trace) equivalent, but not bisimilar.
More precisely: TV-distance is 0, but bisimilarity distance is 1.

[D. Chen, F. van Breugel, J. Worrell, FoSSaCS'12]:
TV-distance $\leq$ bisimilarity distance

Results and Open Problems

Positive Results:

- There is a maximising event.
- The distance can be approximated within arbitrary precision.
- The distance-1 problem is in P.

Negative Results:

- The maximising event may not be ω-regular.
- The threshold problem is NP-hard and hard for square-root-sum.
- The distance is related to Bernoulli convolutions.

Open Questions:

- Efficient approximation?
- Is the threshold problem decidable?

Pisot Number: Definition

A Pisot number is a real algebraic integer greater than 1 such that all its Galois conjugates are less than 1 in absolute value.

Smallest Pisot number (≈ 1.3247): the real root of $x^3 - x - 1$

Another one is the golden ratio $\dfrac{\sqrt{5}+1}{2} \approx 1.6180$.