Linear Distances between Markov Chains

Przemysław Daca[1] Thomas A. Henzinger[1]

Jan Křetínský[2] Tatjana Petrov[1]

[1]IST Austria

[2]Technische Universität München, Germany

CONCUR
Québec, Canada
August 25, 2016

Kinds of properties

classical examples

	logic: model checking	behaviour: equivalence checking
linear	LTL	trace equivalence
branching	CTL	bisimulation

Kinds of properties

probabilistic examples

	logic: model checking	behaviour: equivalence checking
linear	probabilistic LTL	probabilistic trace equivalence
branching	probabilistic CTL	probabilistic bisimulation

Kinds of properties

quantitative probabilistic examples

	logic: model checking	behaviour: equivalence checking
linear	probabilistic LTL	probabilistic trace equivalence
branching	probabilistic CTL	probabilistic bisimulation

Kinds of properties

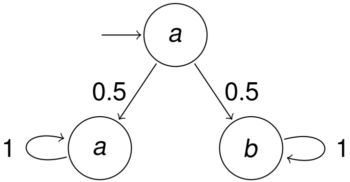
quantitative probabilistic examples

	logic: model checking	behaviour: equivalence checking
linear	probabilistic LTL	probabilistic trace distance
branching	probabilistic CTL	probabilistic bisimulation distance

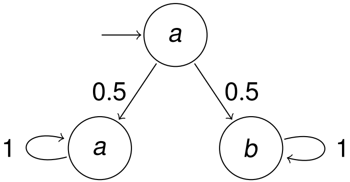
Setting and Outline

1. model: Markov chains
2. properties: linear distances
   - framework
   - examples
3. results for a *statistical (**black-box**)* approach
   - undecidability for total-variation distance
   - decidability for trace distance
   - consequences for other distances

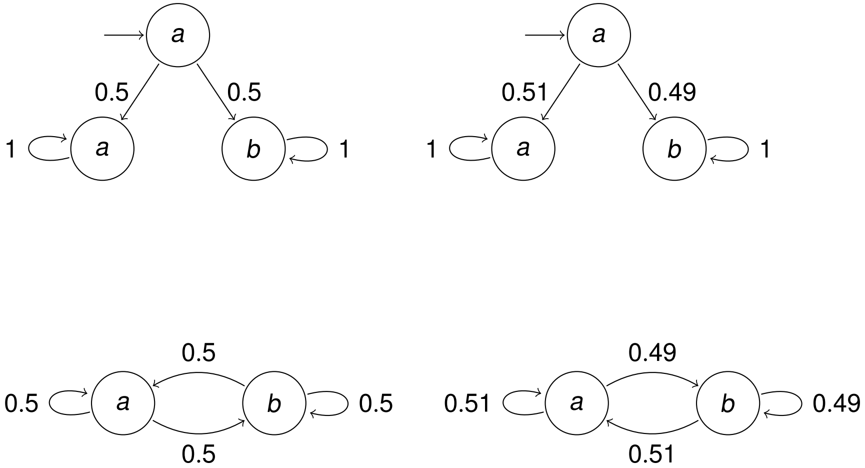
Markov chains



Markov chains



Markov chains



Distances

Definition ($\mathcal{L}$-distance)

For a class $\mathcal{L} \subseteq \mathcal{F}$ of measurable ω-languages, the $\mathcal{L}$-distance $\mathsf{D}_{\mathcal{L}}$ is

$$\mathsf{D}_{\mathcal{L}}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{L \in \mathcal{L}} |\mathbb{P}_1(L) - \mathbb{P}_2(L)|$$

Distances

Definition ($\mathcal{L}$-distance)

For a class $\mathcal{L} \subseteq \mathcal{F}$ of measurable ω-languages, the $\mathcal{L}$-distance $D_{\mathcal{L}}$ is

$$D_{\mathcal{L}}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{L \in \mathcal{L}} |\mathbb{P}_1(L) - \mathbb{P}_2(L)|$$

Example

total variation distance all measurable sets

finite-trace distance generators $w\Sigma^\omega$, where $w \in \Sigma^*$

infinite-trace distance elementary events $w \in \Sigma^\omega$

topological distances clopen sets, Borel hierarchy etc.

automata distances languages recognized by an automata class, e.g. ω-regular

logical distances a logic induces languages of models of its formulae, e.g. LTL, LTL(**X**), LTL(**F**,**G**)

White-box vs. Black-box

- analysis yields precise answers (if decidable)
- complexity grows with state space
- requires full knowledge of the system

- statistics on simulations yields confidence intervals (bounded error)
- complexity depends mostly on the required confidence
- may require some information about the system

White-box vs. Black-box

□

- analysis yields precise answers (if decidable)
- complexity grows with state space
- requires full knowledge of the system

■

- statistics on simulations yields confidence intervals (bounded error)
- complexity depends mostly on the required confidence
- may require some information about the system

Problem

For a given $D_{\mathcal{L}}$, provide an almost-surely terminating algorithm:
Input:

- simulation runs from Markov chains $\mathcal{M}_1$ and $\mathcal{M}_2$
- confidence $\alpha \in (0, 1)$
- interval width $\delta \in (0, 1)$

Output: interval I with $|I| \leq \delta$ and $\mathbb{P}[D_{\mathcal{L}}(\mathcal{M}_1, \mathcal{M}_2) \in I] \geq 1 - \alpha$

White-box vs. Black-box

□

- analysis yields precise answers (if decidable)
- complexity grows with state space
- requires full knowledge of the system

■

- statistics on simulations yields confidence intervals (bounded error)
- complexity depends mostly on the required confidence
- may require some information about the system

Problem

For a given $D_{\mathcal{L}}$, provide an almost-surely terminating algorithm:
Input:

- simulation runs from Markov chains $\mathcal{M}_1$ and $\mathcal{M}_2$
- confidence $\alpha \in (0, 1)$
- interval width $\delta \in (0, 1)$
- lower bound $p_{\min} > 0$ on the minimum transition probability, possibly upper bound n on the size of state spaces (see TACAS'16)

Output: interval I with $|I| \leq \delta$ and $\mathbb{P}[D_{\mathcal{L}}(\mathcal{M}_1, \mathcal{M}_2) \in I] \geq 1 - \alpha$

Results

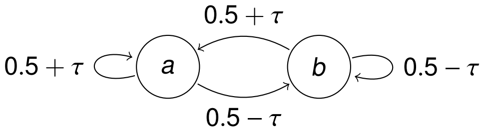
Inestimability of total-variation distance

Total variation distance $\sup_{\text{measurable } L} |\mathbb{P}_1(L) - \mathbb{P}_2(L)|$

Inestimability of total-variation distance

Total variation distance $\sup_{\text{measurable } L} |\mathbb{P}_1(L) - \mathbb{P}_2(L)|$

$\mathcal{M}(\tau)$:



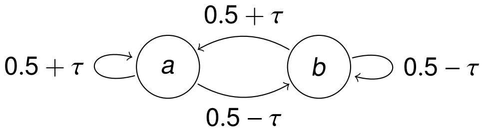
For a fixed τ,
L_n := "there is $\leq (0.5 + \tau/2)n$ symbols a in the path prefix of length n"

$\lim_{n\to\infty} \mathbb{P}_{\mathcal{M}(0)}[L_n] = 1$
$\lim_{n\to\infty} \mathbb{P}_{\mathcal{M}(\tau)}[L_n] = 0$
$\Rightarrow$ total variation distance $= 1$

Inestimability of total-variation distance

Total variation distance $\sup_{\text{measurable } L} |\mathbb{P}_1(L) - \mathbb{P}_2(L)|$

$\mathcal{M}(\tau)$:



For a fixed τ,
L_n := "there is $\leq (0.5 + \tau/2)n$ symbols a in the path prefix of length n"

$\lim_{n\to\infty} \mathbb{P}_{\mathcal{M}(0)}[L_n] = 1$
$\lim_{n\to\infty} \mathbb{P}_{\mathcal{M}(\tau)}[L_n] = 0$
$\Rightarrow$ total variation distance $= 1$

But the necessary simulation run length depends on τ

Estimability of trace distance 1/3

Finite-trace distance $$\mathsf{D}_{\mathsf{FT}}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{w \in \Sigma^*} |\mathbb{P}_1(w\Sigma^\omega) - \mathbb{P}_2(w\Sigma^\omega)|$$

Estimability of trace distance 1/3

Finite-trace distance $$\mathrm{D}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{w \in \Sigma^*} |\mathbb{P}_1(w\Sigma^\omega) - \mathbb{P}_2(w\Sigma^\omega)|$$

The most differentiating trace has unknown length

$$\mathrm{D}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{k \in \mathbb{N}} \mathrm{D}^{k}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2)$$

Estimability of trace distance 1/3

Finite-trace distance $\quad \mathrm{D}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{w \in \Sigma^*} |\mathbb{P}_1(w\Sigma^\omega) - \mathbb{P}_2(w\Sigma^\omega)|$

The most differentiating trace has unknown length

$$\mathrm{D}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{k \in \mathbb{N}} \mathrm{D}^k_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2)$$

1. Computing $\mathrm{D}^k_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2)$ can be done.
2. We can bound the necessary length of the simulation runs!

$$\mathrm{D}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2) = \max_{k < K} \mathrm{D}^k_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2)$$

Estimability of trace distance 1/3

Finite-trace distance $$\mathrm{D}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{w \in \Sigma^*} |\mathbb{P}_1(w\Sigma^\omega) - \mathbb{P}_2(w\Sigma^\omega)|$$

The most differentiating trace has unknown length

$$\mathrm{D}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{k \in \mathbb{N}} \mathrm{D}^{k}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2)$$

1. Computing $\mathrm{D}^{k}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2)$ can be done.
2. We can bound the necessary length of the simulation runs!

$$\mathrm{D}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2) = \max_{k < K} \mathrm{D}^{k}_{\mathrm{FT}}(\mathcal{M}_1, \mathcal{M}_2)$$

$$K = n^2 \left\lceil \frac{\log \epsilon}{\log(1 - p_{\min}^{n^2})} \right\rceil + 2n$$

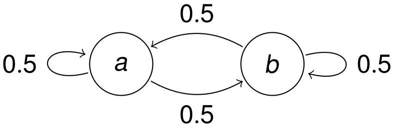
Estimability of trace distance 2/3

Ignore too long traces

Idea:
The longer the trace...

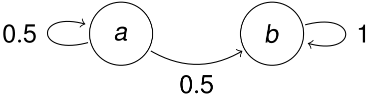
1. either the less probable it is

   ⇒ too long ones can be ignored when *approximating* D_{FT}



2. or its probability stays the same

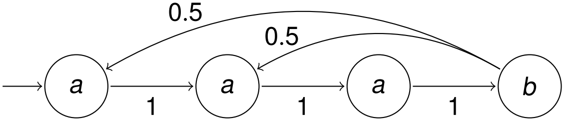
   ⇒ detect these "deterministic" cases early



Estimability of trace distance 3/3

When does the “deterministic” case happen?

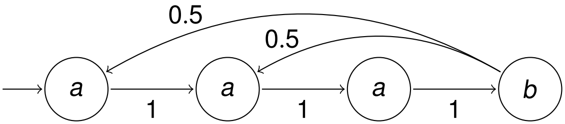
“deterministic” for ... steps ⇒ “deterministic” forever



Estimability of trace distance 3/3

When does the "deterministic" case happen?

"deterministic" for n^2 steps $\Rightarrow$ "deterministic" forever



Consequences:

1. branching case $\Rightarrow$ too long ones can be ignored:

   word of length k $\Rightarrow$ probability $\leq \left(1 - p_{\min}^{n^2}\right)^{\left\lfloor \frac{k}{n^2} \right\rfloor}$

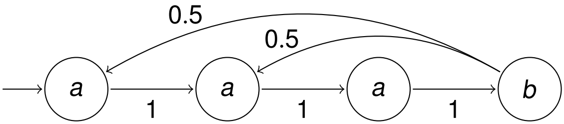
2. "deterministic" case $\Rightarrow$ can be short

   ultimately periodic word

Estimability of trace distance 3/3

When does the "deterministic" case happen?

"deterministic" for n^2 steps $\Rightarrow$ "deterministic" forever



Consequences:

1. branching case $\Rightarrow$ too long ones can be ignored:

   word of length k $\Rightarrow$ probability $\leq \left(1 - p_{\min}^{n^2}\right)^{\left\lfloor \frac{k}{n^2} \right\rfloor}$

2. "deterministic" case $\Rightarrow$ can be short

   ultimately periodic word

3. $$K = n^2 \left\lceil \frac{\log \epsilon}{\log(1 - p_{\min}^{n^2})} \right\rceil + 2n$$

Consequences for other distances

Inestimability

- clopen sets
- LTL($\mathbf{X}, \wedge, \vee$)
- automata

Estimability

- infinite-trace distance
- LTL($\mathbf{FG}, \mathbf{GF}$)
- automata of bounded size
- models with finite precision

Conclusion

Summary

- linear-distance framework for Markov chains
- estimating distances in black-box setting
- (in)estimability border for distances given topologically, logically, and by automata

Future work

- characterize the largest LTL fragment ensuring estimability
- efficient algorithms

Conclusion

Summary

- linear-distance framework for Markov chains
- estimating distances in black-box setting
- (in)estimability border for distances given topologically, logically, and by automata

Future work

- characterize the largest LTL fragment ensuring estimability
- efficient algorithms

Thank you