Linear Distances between Markov Chains

Przemysław Daca¹ Thomas A. Henzinger¹

Jan Křetínský² Tatjana Petrov¹

¹IST Austria
²Technische Universität München, Germany

CONCUR
Québec, Canada
August 25, 2016
Kinds of properties

classical examples

<table>
<thead>
<tr>
<th>logic: model checking</th>
<th>behaviour: equivalence checking</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear branching</td>
<td>LTL</td>
</tr>
<tr>
<td></td>
<td>CTL</td>
</tr>
</tbody>
</table>
## Kinds of properties

<table>
<thead>
<tr>
<th>probabilistic examples</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>logic: model checking</th>
<th>behaviour: equivalence checking</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear branching</td>
<td></td>
</tr>
<tr>
<td>probabilistic LTL</td>
<td>probabilistic trace equivalence</td>
</tr>
<tr>
<td>probabilistic CTL</td>
<td>probabilistic bisimulation</td>
</tr>
</tbody>
</table>
## Kinds of properties

- **Quantitative** probabilistic examples

<table>
<thead>
<tr>
<th>Linear branching</th>
<th>Logic: model checking</th>
<th>Behaviour: equivalence checking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>probabilistic LTL</td>
<td>probabilistic trace equivalence</td>
</tr>
<tr>
<td></td>
<td>probabilistic CTL</td>
<td>probabilistic bisimulation</td>
</tr>
</tbody>
</table>
quantitative probabilistic examples

<table>
<thead>
<tr>
<th></th>
<th>logic: model checking</th>
<th>behaviour: equivalence checking</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>probabilistic LTL</td>
<td>probabilistic trace distance</td>
</tr>
<tr>
<td>branching</td>
<td>probabilistic CTL</td>
<td>probabilistic bisimulation distance</td>
</tr>
</tbody>
</table>
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

\[a\xrightarrow{0.5} a, 0.5 \xrightarrow{a} b, 1 \xrightarrow{b} a, 1 \xrightarrow{b} b\]
Markov chains
Markov chains
Definition ($\mathcal{L}$-distance)

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

$$D_{\mathcal{L}}(M_1, M_2) = \sup_{L \in \mathcal{L}} |\mathbb{P}_1(L) - \mathbb{P}_2(L)|$$
Distances

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

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

$$D_\mathcal{L}(M_1, M_2) = \sup_{L \in \mathcal{L}} |P_1(L) - P_2(L)|$$

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. $\omega$-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

Problem

For a given $D_L$, provide an almost-surely terminating algorithm:

**Input:**
- Simulation runs from Markov chains $M_1$ and $M_2$
- Confidence $\alpha \in (0, 1)$
- Interval width $\delta \in (0, 1)$
- Lower bound $p_{\text{min}} > 0$ 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 $P[D_L(M_1, M_2) \in I] \geq 1 - \alpha$
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 $D_L$, provide an almost-surely terminating algorithm:

Input:
- simulation runs from Markov chains $M_1$ and $M_2$
- confidence $\alpha \in (0, 1)$
- interval width $\delta \in (0, 1)$

Output: interval $I$ with $|I| \leq \delta$ and $\mathbb{P}[D_L(M_1, M_2) \in I] \geq 1 - \alpha$
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 \(D_L\), 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 \(l\) with \(|l| \leq \delta\) and \(\mathbb{P}[D_L(\mathcal{M}_1, \mathcal{M}_2) \in l] \geq 1 - \alpha\)
Results
Total variation distance  \[ \sup_{\text{measurable } L} |P_1(L) - P_2(L)| \]
Inestimability of total-variation distance

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

\[ M(\tau) : \]

\begin{align*}
0.5 + \tau & \quad \rightarrow \\
a & \quad \leftrightarrow \\
0.5 - \tau & \quad \rightarrow \\
b & \quad \leftrightarrow \\
0.5 + \tau & \quad \rightarrow \\
0.5 - \tau &
\end{align*}

For a fixed \( \tau \),
\( L_n := \text{“there is } \leq (0.5 + \tau/2)n \text{ symbols } \mathbf{a} \text{ in the path prefix of length } n\)"

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

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

\[ M(\tau) : \]

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

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

But the necessary simulation run length depends on \( \tau \)
Finite-trace distance

\[ D_{FT}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{w \in \Sigma^*} |P_1(w\Sigma^\omega) - P_2(w\Sigma^\omega)| \]
Estimability of trace distance 1/3

Finite-trace distance

\[ D_{FT}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{w \in \Sigma^*} |\mathbb{P}_1(w \Sigma^\omega) - \mathbb{P}_2(w \Sigma^\omega)| \]

The most differentiating trace has unknown length

\[ D_{FT}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{k \in \mathbb{N}} D^k_{FT}(\mathcal{M}_1, \mathcal{M}_2) \]
Finite-trace distance
\[ D_{FT}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{w \in \Sigma^*} |\mathbb{P}_1(w^\omega) - \mathbb{P}_2(w^\omega)| \]

The most differentiating trace has unknown length

\[ D_{FT}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{k \in \mathbb{N}} D_{FT}^k(\mathcal{M}_1, \mathcal{M}_2) \]

1. Computing \( D_{FT}^k(\mathcal{M}_1, \mathcal{M}_2) \) can be done.

2. We can bound the necessary length of the simulation runs!

\[ D_{FT}(\mathcal{M}_1, \mathcal{M}_2) = \max_{k<K} D_{FT}^k(\mathcal{M}_1, \mathcal{M}_2) \]
Finite-trace distance

\[ D_{FT}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{w \in \Sigma^*} |\mathbb{P}_1(w\Sigma^\omega) - \mathbb{P}_2(w\Sigma^\omega)| \]

The most differentiating trace has unknown length

\[ D_{FT}(\mathcal{M}_1, \mathcal{M}_2) = \sup_{k \in \mathbb{N}} D_{FT}^k(\mathcal{M}_1, \mathcal{M}_2) \]

1. Computing \( D_{FT}^k(\mathcal{M}_1, \mathcal{M}_2) \) can be done.

2. We can bound the necessary length of the simulation runs!

\[ D_{FT}(\mathcal{M}_1, \mathcal{M}_2) = \max_{k < K} D_{FT}^k(\mathcal{M}_1, \mathcal{M}_2) \]

\[ K = n^2 \left\lfloor \frac{\log \epsilon}{\log(1 - p_{\min}^{n^2})} \right\rfloor + 2n \]
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
When does the “deterministic” case happen?

“deterministic” for ... steps ⇒ “deterministic” forever
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
When does the “deterministic” case happen?

“deterministic” for \( n^2 \) steps ⇒ “deterministic” forever

Consequences:

1. branching case ⇒ too long ones can be ignored:
   word of length \( k \) ⇒ probability \( \leq (1 - p_{\text{min}}^{n^2})^{\lfloor k/n^2 \rfloor} \)

2. “deterministic” case ⇒ can be short
   ultimately periodic word

3. 
   \[
   K = n^2 \left( \frac{\log \epsilon}{\log(1 - p_{\text{min}}^{n^2})} \right) + 2n
   \]
Consequences for other distances

**Inestimability**
- clopen sets
- LTL(\(\textbf{X}, \land, \lor\))
- automata

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