Probabilistic Aspects of Computer Science: Probabilistic Automata

Serge Haddad

LSV, ENS Paris-Saclay & CNRS & Inria

M1 Jacques Herbrand

1 Presentation

2 Properties of Stochastic Languages

3 Decidability Results
Plan

1. Presentation

Properties of Stochastic Languages

Decidability Results
An introductive example

Planning holidays in a foreign country

1. Choosing which plane company to use: lowcost or highcost;
2. Renting a room in an hotel by internet or phone;
3. Buying tickets for some exhibitions with agency: seeall or dontmiss.

Usually these actions must be planned before the holidays.

Thus one looks for an \textit{a priori} optimal policy that maximizes the probability to \textit{reach} a goal.
The probability of success of lowcost · internet · seeall is $\frac{27}{64}$. 
Probabilistic Automata (PA) are a variation of MDP where:

- The set of possible actions is the same in every state.
- There are no rewards.
- There is a subset of final states.

More formally, a PA $A = (Q, A, \{P_a\}_{a \in A}, \pi_0, F)$ is defined by:

- $Q$, the finite set of states;
- $A$, the finite alphabet;
- For all $a \in A$, $P_a$, a probability transition matrix over $S$;
- $\pi_0$, the initial distribution over states and $F \subseteq Q$ the final states.
An edge from a state to another one is labelled by a vector of transition probabilities indexed by $A$. The vector is denoted by a formal sum. For instance, the transition from $q_0$ to itself is labelled by $1a + 0.5b$ means that:

- when $a$ is chosen in state $q_0$, the probability that the next state is $q_0$, $P_a[q_0, q_0]$, is equal to 1.
- when $b$ is chosen in state $q_0$, the probability that the next state is $q_0$, $P_b[q_0, q_0]$, is equal to 0.5.

$A = \{a, b\}$;

$Q = \{q_0, q_1\}$, $F = \{q_1\}$;

$\pi_0[q_0] = 1$. 

Illustration

![Diagram](image-url)
Policies in PA

Words are policies. When some finite word \( w \overset{\text{def}}{=} a_1 \ldots a_n \) is selected, we are interested in the probability to be in a final state using \( w \) as a policy.

Given \( \mathcal{A} \) a PA and \( w \overset{\text{def}}{=} a_1 \ldots a_n \in A^* \) a word, the acceptance probability of \( w \) by \( \mathcal{A} \) is defined by:

\[
\Pr_{\mathcal{A}}(w) \overset{\text{def}}{=} \sum_{q \in Q} \pi_0[q] \sum_{q' \in F} \left( \prod_{i=1}^n P_{a_i} \right) [q, q']
\]

Notation. Given a word \( w \overset{\text{def}}{=} a_1 \ldots a_n \), the probability matrix \( P_w \) is defined by \( P_w \overset{\text{def}}{=} \prod_{i=1}^n P_{a_i} \). In particular \( P_\varepsilon = \text{Id} \).

With these notations:

\[
\Pr_{\mathcal{A}}(w) = \pi_0 P_w 1^T_F
\]

where \( 1_F \) is the indicator vector of subset \( F \).
Computation of \( \text{Pr}_A(abba) \) by induction w.r.t. the prefixes. First \( \text{Pr}_A(\varepsilon) = 0 \).

- \( \text{Pr}_A(a) = \frac{1}{2} \text{Pr}_A(\varepsilon) = 0 \)
- \( \text{Pr}_A(ab) = \text{Pr}_A(a) + \frac{1}{2}(1 - \text{Pr}_A(a)) = \frac{1}{2} \)
- \( \text{Pr}_A(abb) = \text{Pr}_A(ab) + \frac{1}{2}(1 - \text{Pr}_A(ab)) = \frac{3}{4} \)
- \( \text{Pr}_A(abba) = \frac{1}{2} \text{Pr}_A(abb) = \frac{3}{8} \)

More generally, the following recursive equations hold:

\[
\text{Pr}_A(wa) = \frac{1}{2} \text{Pr}_A(w) \quad \text{and} \quad \text{Pr}_A(wb) = \frac{1}{2}(1 + \text{Pr}_A(w))
\]

from which one can derive an explicit expression of the acceptance probability:

\[
\text{Pr}_A(a_1 \ldots a_n) = \sum_{i=1}^{n} 2^{i-n-1} \cdot 1_{a_i=b}
\]

Which word maximizes the acceptance probability?
Stochastic languages

We are interested in “useful” policies.

This directly leads to the introduction of stochastic languages. Let:

- $\mathcal{A}$ be a probabilistic automaton;
- $\theta \in [0, 1]$ be a threshold also called a cut point;
- $\bowtie \in \{<, \leq, >, \geq, =, \neq\}$ be a comparison operator.

Then $L_{\bowtie \theta}(\mathcal{A})$ is defined by:

$$L_{\bowtie \theta}(\mathcal{A}) = \{w \in A^* | \Pr_{\mathcal{A}}(w) \bowtie \theta\}$$

For expressiveness and decidability issues, one also needs the following definitions.

- A rational PA is a PA with probability distributions over $\mathbb{Q}^Q$.
- A rational stochastic language is a stochastic language specified by a rational PA and a rational threshold.
Counting with \( \text{PA} \)

(a succinct representation with an omitted absorbing rejecting state)

Any word \( z \) different from \( a^m b^n \) with \( m > 0, n > 0 \) cannot be accepted.

Let \( w \overset{\text{def}}{=} a^m b^n \) with \( m > 0, n > 0 \). \( w \) can be accepted by:

- a path \( q_0, q_1^m, q_2^n \) with probability \( \frac{1}{2^n} \);
- or by a family of paths \( q_0, q_3^r, q_4^s, q_5^n \) with \( 0 < r, s \) and \( r + s = m \) with cumulated probability \( \frac{1}{2} - \frac{1}{2^m} \).

Summing, one obtains: \( \frac{1}{2} + \frac{1}{2^n} - \frac{1}{2^m} \).

Thus: \( \mathcal{L}_{=0.5}(\mathcal{A}) = \{ a^n b^n \mid n > 0 \} \)
Plan

Presentation

2 Properties of Stochastic Languages

Decidability Results
Expressiveness problems

Provide a minimal set of comparison operators and thresholds.

Position the stochastic languages w.r.t. the Chomsky hierarchy.

Study the closure properties of the stochastic languages.
A single threshold is enough

\begin{align*}
\mathcal{A} & \xleftarrow{\alpha \pi_0[q]} q \\
q_0 & \xrightarrow{1 - \alpha} 1A
\end{align*}

The value \( \alpha \) depends on \( \theta \neq \frac{1}{2} \) in the following way:

- If \( \theta > \frac{1}{2} \) then \( q_0 \notin F \) and \( \alpha \overset{\text{def}}{=} \frac{1}{2\theta} \) so that for all \( w \in A^* \),
  \[ \Pr_{\mathcal{A}'}(w) = \frac{1}{2\theta} \Pr_{\mathcal{A}}(w) \]
  Thus \( w \in L_{\bowtie\theta}^{\frac{1}{2}}(\mathcal{A}') \) iff \( w \in L_{\bowtie\theta}(\mathcal{A}) \).

- If \( \theta < \frac{1}{2} \) then \( q_0 \in F \) and \( \alpha \overset{\text{def}}{=} \frac{1}{2(1 - \theta)} \) so that for all \( w \in A^* \),
  \[ \Pr_{\mathcal{A}'}(w) = \frac{1 - 2\theta + \Pr_{\mathcal{A}}(w)}{2(1 - \theta)} \]
  Thus \( w \in L_{\bowtie\theta}^{\frac{1}{2}}(\mathcal{A}') \) iff \( w \in L_{\bowtie\theta}(\mathcal{A}) \).
Getting rid of (dis)equality

Given a PA $A$, we build $A'$ as follows.

- The set of states $Q' \overset{\text{def}}{=} Q \times Q$;
- $P'_a[(q_1, q_2), (q'_1, q'_2)] \overset{\text{def}}{=} P_a[q_1, q'_1] P_a[q_2, q'_2]$;
- $\pi'_0[q_1, q_2] \overset{\text{def}}{=} \pi_0[q_1] \pi_0[q_2]$ and $F' \overset{\text{def}}{=} F \times (Q \setminus F)$.

Once a word $w$ is selected, the two components of the DES behave independently and so:

$$Pr_{A'}(w) = Pr_{A}(w)(1 - Pr_{A}(w))$$

Consequently $Pr_{A'}(w) \leq \frac{1}{4}$ with equality iff $Pr_{A}(w) = \frac{1}{2}$. Thus:

$$L_{\geq \frac{1}{4}}(A') = L_{= \frac{1}{2}}(A)$$
Getting rid of “lower (or equal) than”

Given a PA $\mathcal{A}$, we build $\mathcal{A}'$ by complementing the final states. Then:

$$\Pr_{\mathcal{A}'}(w) = 1 - \Pr_{\mathcal{A}}(w)$$

And so:

$$L_{\geq \theta}({\mathcal{A}'}) = L_{< \theta}(\mathcal{A})$$

$$L_{> \theta}({\mathcal{A}'}) = L_{\leq \theta}(\mathcal{A})$$
### The Chomsky hierarchy

<table>
<thead>
<tr>
<th>Class</th>
<th>Grammar</th>
<th>Device</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular language</td>
<td>( L \to aR</td>
<td>a</td>
</tr>
<tr>
<td>Algebraic language</td>
<td>( L \to R_1 \ldots R_n ) with ( L \in \Delta ) and ( R_i \in \Delta \cup \Sigma )</td>
<td>Stack automaton</td>
</tr>
<tr>
<td>Context-sensitive language</td>
<td>( L_1 \ldots L_m \to R_1 \ldots R_n ) with ( m \leq n, (S \to \varepsilon) ) with ( L_i, R_j \in \Delta \cup \Sigma )</td>
<td>Non determ. Turing machine performing in linear space</td>
</tr>
<tr>
<td>Recursively enumerable language</td>
<td>( L_1 \ldots L_m \to R_1 \ldots R_n ) avec ( L_i, R_j \in \Delta \cup \Sigma )</td>
<td>Turing machine</td>
</tr>
</tbody>
</table>
Revisiting the Chomsky hierarchy

- recursively enumerable
- context sensitive
- algebraic
- regular
- rational stochastic
- stochastic
Define \( v_a \equiv 0 \) and \( v_b \equiv 1 \).

The acceptance probability of \( w_1 \ldots w_n \) is the binary number \( 0.v_{w_n} \ldots v_{w_1} \).

So \( \mathcal{L}_{>\theta}(\mathcal{A}) \) is the set of representations of numbers (with finite binary development) greater than \( \theta \).

Thus given \( 0 \leq \theta < \theta' \leq 1 \),

\[
\mathcal{L}_{>\theta'}(\mathcal{A}) \subsetneq \mathcal{L}_{>\theta}(\mathcal{A})
\]

So there is an uncountable number of stochastic languages implying that “most” of them are non recursively enumerable.

This result does not hold for rational stochastic languages.
A deterministic automaton is a stochastic automaton with probabilities in \( \{0, 1\} \).

Thus regular languages are stochastic languages.

The language \( \{a^n b^n \mid n > 0\} \) is a rational stochastic non regular language.
Non stochastic context-free languages (1)

\[ L \overset{\text{def}}{=} \{ a^{n_1}ba^{n_2}b\ldots a^{n_k}b a^* \mid \exists i > 1 \; n_i = n_1 \} \]

is a non stochastic context-free language.

Proof.

\( L \) is context-free. Use a non deterministic automaton with a counter.

- With a counter one counts \( n_1 \) the number of \( a \)'s until the first occurrence of \( b \).
- Then one guesses an occurrence of \( b \) and decrements the counter by the occurrences of \( a \) until the next occurrence of \( b \).
- If the counter is zero the word is accepted.

Assume that (1) \( L = L_{>\theta}(A) \) or (2) \( L = L_{\geq \theta}(A) \).

Let \( \sum_{i=0}^{n} c_i x^i \) be the minimal polynomial of \( P_a \).

Since 1 is an eigenvalue of \( P_a \), one gets \( \sum_{i=0}^{n} c_i = 0 \) and there are positive and negative coefficients.

By definition, \( \sum_{i=0}^{n} c_i P_a^i = 0 \) and so for any word \( w \), \( \sum_{i=0}^{n} c_i P_a^i w = 0 \).
Non stochastic context-free languages (2)

Proof (continued).

Let $Pos = \{ i \mid 0 \leq i \leq n \land c_i > 0 \}$ and $NonPos = \{ i \mid 0 \leq i \leq n \land c_i \leq 0 \}$.

Write $Pos$ as $\{ i_1, \ldots, i_k \}$.

Choose $w \overset{\text{def}}{=} ba^{i_1}b \ldots ba^{i_k}b$.

**Case** $L = L_{>\theta}(A)$. Let $0 \leq i \leq n$, by definition of $L$,

$$
\pi_0 P_{a^i w} 1^T_F > \theta \text{ iff } i \in \{ i_1, \ldots, i_k \}
$$

So:

$$
0 = \sum_{i=0}^n c_i \pi_0 P_{a^i w} 1^T_F = \sum_{i \in Pos} c_i \pi_0 P_{a^i w} 1^T_F + \sum_{i \in NonPos} c_i \pi_0 P_{a^i w} 1^T_F > (\sum_{i \in Pos} c_i) \theta + (\sum_{i \in NonPos} c_i) \theta = (\sum_{i=0}^n c_i) \theta = 0
$$

leading to a contradiction.

**Case** $L = L_{\geq \theta}(A)$. Let $0 \leq i \leq n$, by definition of $L$,

$$
\pi_0 P_{a^i w} 1^T_F \geq \theta \text{ iff } i \in \{ i_1, \ldots, i_k \}
$$

So:

$$
0 = \sum_{i=0}^n c_i \pi_0 P_{a^i w} 1^T_F = \sum_{i \in Pos} c_i \pi_0 P_{a^i w} 1^T_F + \sum_{i \in NonPos} c_i \pi_0 P_{a^i w} 1^T_F > (\sum_{i \in Pos} c_i) \theta + (\sum_{i \in NonPos} c_i) \theta = (\sum_{i=0}^n c_i) \theta = 0
$$

leading to a contradiction.
Non context-free stochastic languages (1)

\[ L \overset{\text{def}}{=} \{ a^n b^n c^n \mid n > 0 \} \]

is a non context-free rational stochastic language.

**Proof.**

Using Ogden’s lemma it can be easily proved that \( L \) is not context-free.

Observe that \( L = L_1 \cap L_2 \) with \( L_1 \overset{\text{def}}{=} \{ a^n b^n c^+ \mid n > 0 \} \) and \( L_2 \overset{\text{def}}{=} \{ a^+ b^n c^n \mid n > 0 \} \).

So we prove that:

- for \( i \in \{1, 2\} \), \( L_i = L_{=\frac{1}{2}}(\mathcal{A}_i) \) for some \( \mathcal{A}_i \)
- the family of languages \( \{ L = L_{=\frac{1}{2}}(A) \} \) is closed under intersection.
Proof (continued).

\[ L_{\frac{1}{2}}(A) = \{a^n b^n c^+ \mid n > 0\} \]
Non context-free stochastic languages (3)

Proof (ended).
Let \( L = \frac{1}{2}(A_1) \) and \( L = \frac{1}{2}(A_2) \) be two arbitrary languages.
Using the previous construction, let \( A'_1 \) and \( A'_2 \) be automata such that:

- For any word \( w \), \( \Pr_{A'_i}(w) \leq \frac{1}{4} \);
- \( L = \frac{1}{2}(A_i) = L = \frac{1}{4}(A'_i) \).

One builds \( A \) as follows:

- The set of states \( Q \overset{\text{def}}{=} Q'_1 \times Q'_2 \);
- \( P_a[(q_1, q_2), (q'_1, q'_2)] \overset{\text{def}}{=} (P'_1)_a[q_1, q'_1](P'_2)_a[q_2, q'_2] \);
- \( \pi'_0[q_1, q_2] \overset{\text{def}}{=} \pi_{1,0}[q_1] \pi_{2,0}[q_2] \) and \( F \overset{\text{def}}{=} F'_1 \times F'_2 \).

By construction, \( \Pr_A(w) = \Pr_{A'_1}(w)\Pr_{A'_2}(w) \).
So for all word \( w \), \( \Pr_A(w) \leq \frac{1}{16} \) and \( \Pr_A(w) = \frac{1}{16} \) iff \( \Pr_{A'_1}(w) = \Pr_{A'_2}(w) = \frac{1}{4} \).

Consequently,
\[
L = \frac{1}{16}(A) = L = \frac{1}{2}(A_1) \cap L = \frac{1}{2}(A_2)
\]
Inclusion in context-sensitive languages

The class of rational stochastic languages is strictly included in the class of context-sensitive languages.

Proof.

Context-sensitive languages are the languages for which membership checking can be performed by a non deterministic procedure in linear space.

**A deterministic procedure in linear space (far from being optimal)**

Pre-computation in constant space.

- Compute the l.c.m., say $b$, of denominators of $\theta$, items of matrices $\{P_a\}_{a \in A}$ and, items of vector $\pi_0$.
- Build the integer matrices $P'_a \overset{\text{def}}{=} bP_a$ and vector $\pi'_0 \overset{\text{def}}{=} b\pi_0$.

For word $w \overset{\text{def}}{=} a_1 \ldots a_n$, the problem becomes

$$\pi'_0 (\prod_{i=1}^n P'_{a_i}) 1_T^T \bowtie \theta b^{n+1}?$$

- Compute $\theta b^{n+1}$ in space $O(n)$.
- Compute $v \overset{\text{def}}{=} \pi'_0 (\prod_{i=1}^n P'_{a_i})$
  by initializing $v$ to $\pi'_0$ and then iteratively multiply it by $P'_{a_i}$.
  The vectors are bounded by $b^{n+1}$. So this is performed in space $O(n)$.
- The sum and comparison are also done in space $O(n)$. 
Operations with regular languages

The family of (rational) stochastic languages is closed under intersection and union with regular languages.

Proof.
Let \( L_{\bowtie \theta}(A_1) \) be a (rational) stochastic language (with \( \bowtie \in \{>, \geq\} \)) and \( L_{=1}(A_2) \) be a regular language.

\[
L_{\bowtie \frac{1}{2}}(A) = L_{\bowtie \theta}(A_1) \cup L_{=1}(A_2) \quad \text{and} \quad L_{\bowtie 1+\frac{1}{2}}(A) = L_{\bowtie \theta}(A_1) \cap L_{=1}(A_2)
\]
A stochastic language

$L = \frac{1}{2} (A) = \{ a^{m_1} b ... b a^{m_k} b \mid 1 < k \land m_1 = m_k \}$

since $\Pr_A(a^{m_1} b ... b a^{m_k} b) = \frac{1}{2} \left( \left( \frac{1}{2} \right)^{k+m_k-1} + 1 - \left( \frac{1}{2} \right)^{k+m_1-1} \right)$
The family of (rational) stochastic languages is not closed under concatenation with a regular language.

Proof.

Let \( L \overset{\text{def}}{=} \{a^{m_1}b \ldots ba^{m_k}b \mid 1 < k \land m_1 = m_k\} \)
be the previous stochastic language.

Then \( LA^* = \{a^{m_1}ba^{m_2}b \ldots a^{m_k}ba^* \mid \exists i > 1 \ m_i = m_1\} \)
which is not a stochastic language.
The family of (rational) stochastic languages is not closed under Kleene star.

**Proof.**

Let \( L \overset{\text{def}}{=} \{ a^m b \ldots b a^m k b \mid 1 < k \land m_1 = m_k \} \) be the previous stochastic language. Assume that \( L^* = L\_\text{\$\theta(\mathcal{A})} \) with \( \_\text{\$\in \{>, \geq\} \).

Let \( \sum_{i=0}^{n} c_i x^i \) be the minimal polynomial of \( P_a \).

Since 1 is an eigenvalue of \( P_a \), one gets \( \sum_{i=0}^{n} c_i = 0 \) and there are positive and negative coefficients.

By definition, \( \sum_{i=0}^{n} c_i P_a^i = 0 \) and so for any word \( w \), \( \sum_{i=0}^{n} c_i P_a^i w = 0 \).

Let \( c_{i_1}, \ldots, c_{i_k} \) be the positive coefficients of this polynomial.

Let \( w \overset{\text{def}}{=} ba^{i_1} b(a^{i_2} b)^2 \ldots (a^{i_k} b)^2 \).

\( a^i w \in L^* \) iff \( i \in \{i_1, \ldots, i_k\} \).

**Case** \( L^* = L\_\text{\$\theta(\mathcal{A})} \).

Let \( 0 \leq i \leq n, P_{a^i w} 1_F^T > \theta \) iff \( i \in \{i_1, \ldots, i_k\} \).

So: \( 0 = \sum_{i=0}^{n} c_i \pi_0 P_{a^i w} 1_F^T > (\sum_{i=0}^{n} c_i) \theta = 0 \)

leading to a contradiction.

**Case** \( L^* = L\_\text{\$\geq\theta(\mathcal{A})} \).

Let \( 0 \leq i \leq n, P_{a^i w} 1_F^T \geq \theta \) iff \( i \in \{i_1, \ldots, i_k\} \).

So: \( 0 = \sum_{i=0}^{n} c_i \pi_0 P_{a^i w} 1_F^T > (\sum_{i=0}^{n} c_i) \theta = 0 \)

leading to a contradiction.
A stochastic language

$L = \frac{1}{2} (A) = \{a^{m_1} b \ldots ba^{m_k} bc A^* \mid 1 < k \land m_1 = m_k\}$
Homomorphism

The family of (rational) stochastic languages is not closed under homomorphism.

Proof.
Let $L \overset{\text{def}}{=} \{ a^{m_1} b \ldots ba^{m_k} bcA^* \mid 1 < k \land m_1 = m_k \}$ be the previous stochastic language.

Define the homomorphism $h$ from $A$ to $A' \overset{\text{def}}{=} \{ a, b \}$ by:

$$h(a) \overset{\text{def}}{=} a \quad h(b) \overset{\text{def}}{=} b \quad h(c) \overset{\text{def}}{=} \varepsilon$$

Then $h(L) = \{ a^{m_1} b a^{m_2} b \ldots a^{m_k} ba^* \mid \exists i > 1 \ m_i = m_1 \}$ which is not a stochastic language.
Plan

Presentation

Properties of Stochastic Languages

3 Decidability Results
Two decision problems

Let $\mathcal{A}$ and $\mathcal{A}'$ be probabilistic automata.

First problem

Are $\mathcal{A}$ and $\mathcal{A}'$ equivalent?

\[ \forall w \in A^* \quad \Pr_{\mathcal{A}}(w) = \Pr_{\mathcal{A}'}(w) \]

Second problem

Is $L_{\triangleright \triangleleft \theta}(\mathcal{A})$ equal to $L_{\triangleright \triangleleft \theta'}(\mathcal{A}')$?

For deterministic automata this is the same problem.
It can be solved in polynomial time by a product construction which provides a witness of non equivalence of size less than $|Q||Q'|$. 
Linear algebra recalls

Let $v_0 \in \mathbb{R}^n$ and $v_1, \ldots, v_k$ be linearly independent vectors of $\mathbb{R}^n$.

How to check whether $v_0$ is a linear combination of $v_1, \ldots, v_k$?

- Solve in $O(k^3 + n^2)$

$$
\begin{pmatrix}
  v_1[1] & \ldots & v_k[1] \\
  \vdots & \ddots & \vdots \\
  v_1[n] & \ldots & v_k[n]
\end{pmatrix}
\begin{pmatrix}
  x_1 \\
  \vdots \\
  x_k
\end{pmatrix}
= 
\begin{pmatrix}
  v_0[1] \\
  \vdots \\
  v_0[n]
\end{pmatrix}
$$

- When $v_1, \ldots, v_k$ are orthogonal
  (i.e. for all $a \neq b$, $v_a \cdot v_b \overset{\text{def}}{=} \sum_{i=1}^{n} v_a[i]v_b[i] = 0$)

Compute in $O(kn)$ the orthogonal projection

$$
w_0 = \sum_{i=1}^{k} \frac{v_0 \cdot v_i}{v_i \cdot v_i} v_i
$$

Check in $O(n)$ whether $v_0 = w_0$. 
Principles of equivalence checking

Enumeration of words
Looking for a counter-example whose length is increasing starting with word $\varepsilon$.

A stack
Managing a stack of words $w$ in order to find counter-examples $aw$ for all $a \in A$. For efficiency purposes, the stack contains tuples $(P_w 1_F, P'_w 1_{F'}, w)$.

An orthogonal family for restricting the enumeration
$Gen$ is a set of independent orthogonal vectors of $\mathbb{R}^{Q\cup Q'}$.
If $w$ is not a counter-example, check if $v \overset{\text{def}}{=} (P_w 1_F, P'_w 1_{F'})$ is generated by $Gen$.

- producing $v'$ the orthogonal projection of $v$ on subspace spanned by $Gen$;
- comparing $v'$ to $v$.

If $v' \neq v$ then:

- $w$ is added to the stack;
- $v - v'$ is added to $Gen$. 
The algorithm

If $\pi_0 \cdot 1_F \neq \pi_0' \cdot 1_{F'}$ then return($false, \varepsilon$)

$Gen \leftarrow \{(1_F, 1_{F'})\}; \text{Push}(Stack, (1_F, 1_{F'}, \varepsilon))$

Repeat

$(v, v', w) \leftarrow \text{Pop}(Stack)$

For $a \in A$ do

$z \leftarrow P_a v; \ z' \leftarrow P_a v'$

If $\pi_0 \cdot z \neq \pi_0' \cdot z'$ then return($false, aw$)

$y \leftarrow 0; \ y' \leftarrow 0$

For $(x, x') \in Gen$ do $(y, y') \leftarrow (y, y') + \frac{z \cdot x + z' \cdot x'}{x \cdot x + x' \cdot x'}(x, x')$

If $(z, z') \neq (y, y')$ then

$\text{Push}(Stack, (z, z', aw))$

$Gen \leftarrow Gen \cup \{(z - y, z' - y')\}$

Until IsEmpty($Stack$)

return(true)
Complexity

Time complexity

An item is pushed on the stack iff an item is added to $Gen$.

There can be no more than $|Q| + |Q'|$ items in $Gen$.

So there are at most $|Q| + |Q'|$ iterations of the external loop.

The index of the first inner loop ranges over $A$ while the index of the most inner loop ranges over $Gen$.

The operations inside the most inner loop are done in $O(|Q| + |Q'|)$.

This leads to an overall time complexity of $O(((|Q| + |Q'|)^3 |A|)$.

Length of witnesses

In addition, the length of the witness is at most $|Q| + |Q'|$.

(Also valid for deterministic automata)
Correctness

Assume that the automata are not equivalent and that the algorithm returns \textbf{true}. Let \( u \) be a non examined word such that \( \Pr_A(u) \neq \Pr_{A'}(u) \).

Let \( u \overset{\text{def}}{=} w'w \) with \( w(\neq u) \) the greatest suffix examined by the algorithm.

Among such words \( u \), pick one word such that \( |w'| \) is minimal.

**Claim.** There exists \( w'' \) that has been inserted in the stack before \( w \) such that \( \Pr_A(w'w'') \neq \Pr_{A'}(w'w'') \).

Let \( \text{Gen} = \{w_1, \ldots, w_k\} \) when examining \( w \), there exist \( \lambda_1, \ldots, \lambda_k \) such that:

\[
\text{So: } P_w 1_F = \sum_{i=1}^{k} \lambda_i P_{w_i} 1_F \quad \text{and} \quad P'_w 1_{F'} = \sum_{i=1}^{k} \lambda_i P'_{w_i} 1_{F'}
\]

\[
\Pr_A(w'w) \overset{\text{def}}{=} \pi_0 P_{w'} P_w 1_F = \sum_{i=1}^{k} \lambda_i \pi_0 P_{w'} P_{w_i} 1_F = \sum_{i=1}^{k} \lambda_i \Pr_A(w'w_i)
\]

Similarly: \( \Pr_{A'}(w'w) = \sum_{i=1}^{k} \lambda_i \Pr_{A'}(w'w_i) \)

So there exists \( i \), with \( \Pr_A(w'w_i) \neq \Pr_{A'}(w'w_i) \).

Let \( w' \overset{\text{def}}{=} w'''a \). \( aw_i \) is examined by the algorithm.

So the word \( u' \overset{\text{def}}{=} w'w_i \) has a decomposition \( u' \overset{\text{def}}{=} z'z \) where \( z \) the greatest suffix examined by the algorithm has for suffix \( aw_i \). So \( |z'| < |w'| \): a contradiction.
Undecidability of the equality problem

Given $\mathcal{A}$ a rational stochastic automaton, the question $L_{\frac{1}{2}}(\mathcal{A}) = \{\varepsilon\}$? is undecidable.

**Proof.**

By reduction of the undecidable Post correspondence problem (PCP):

Given an alphabet $A$ and two morphisms $\varphi_1, \varphi_2$ from $A$ to $\{0, 1\}^+$, does there exist a word $w \in A^+$ such that $\varphi_1(w) = \varphi_2(w)$?

Already undecidable for a restriction where the images of letters lie in $(10 + 11)^+$. Inserting a 1 before each letter of images reduces the former problem to the latter.

A word $w \overset{\text{def}}{=} a_1 \ldots a_n \in (10 + 11)^+$ defines a value $\text{val}(w) \in [0, 1]$ by:

$$\text{val}(w) \overset{\text{def}}{=} \sum_{i=1}^{n} \frac{a_i}{2^{n+1-i}}$$

Since every word starts with a 1, $\text{val}(w) = \text{val}(w')$ implies $w = w'$. 
Reduction of PCP

For $w \in A^+$ and $i \in \{1, 2\}$, define $val_i(w) \overset{\text{def}}{=} val(\varphi_i(w))$.

\[ \sum_a (1 - val_1(a))a \]

\[ \sum_a (1 - val_1(a) - 2^{-|\varphi_1(a)|})a \]

\[ \sum_a val_1(a)a \]

\[ \sum_a (val_1(a) + 2^{-|\varphi_1(a)|})a \]

\[ \sum_a (1 - val_2(a))a \]

\[ \sum_a (1 - val_2(a) - 2^{-|\varphi_2(a)|})a \]

\[ \sum_a val_2(a)a \]

\[ \sum_a (val_2(a) + 2^{-|\varphi_2(a)|})a \]
**Illustration of the reduction**

<table>
<thead>
<tr>
<th>( A )</th>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varphi_1 )</td>
<td>(1)0(1)1</td>
<td>(1)0(1)0</td>
<td>(1)1</td>
</tr>
<tr>
<td>( \varphi_2 )</td>
<td>(1)0</td>
<td>(1)0</td>
<td>(1)1(1)1(1)1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( A )</th>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{val}_1 )</td>
<td>( \frac{13}{16} )</td>
<td>( \frac{7}{16} )</td>
<td>( \frac{3}{4} )</td>
</tr>
<tr>
<td>( \text{val}_2 )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{63}{64} )</td>
</tr>
</tbody>
</table>

\[
\frac{3}{16}a + \frac{9}{16}b + \frac{1}{4}c \quad q_{10} \quad \frac{13}{16}a + \frac{7}{16}b + \frac{3}{4}c \quad q_{11} \quad \frac{7}{8}a + \frac{1}{2}b + c
\]

\[
\frac{1}{8}a + \frac{1}{2}b
\]

\[
\frac{3}{4}a + \frac{3}{4}b + \frac{1}{64}c \quad q_{20} \quad \frac{1}{4}a + \frac{1}{4}b + \frac{63}{64}c \quad q_{21} \quad \frac{1}{2}a + \frac{1}{2}b + c
\]

\[
\frac{1}{2}a + \frac{1}{2}b
\]
Correctness of the reduction

The recurrence equation:

\[ 1_{q_0} P_w a 1^T_{q_1} = 1_{q_0} P_w 1^T_{q_1} (val_i(a) + 2^{-|\varphi_i(a)|}) + (1 - 1_{q_0} P_w 1^T_{q_1})val_i(a) \]

\[ = val_i(a) + 2^{-|\varphi_i(a)|} 1_{q_0} P_w 1^T_{q_1} \]

By induction we obtain that for all \( w \stackrel{\text{def}}{=} a_1 \ldots a_n \):

\[ 1_{q_0} P_w 1^T_{q_1} = \sum_{j=1}^{n} val_i(a_j) 2^{-\sum_{j<k\leq n} |\varphi_i(a_k)|} = val_i(w) \]

So for \( w \in A^+ \): \( \Pr_A(w) = \frac{1}{2}(val_1(w) + 1 - val_2(w)) \).

Thus \( w \in L_{=\frac{1}{2}}(A) \) iff \( val(\varphi_1(w)) = val(\varphi_2(w)) \) implying that \( \varphi_1(w) = \varphi_2(w) \).