On the Memory Consumption of Probabilistic Pushdown Automata

Tomáš Brázdil[1] Javier Esparza[2] Stefan Kiefer[3]

[1]Masaryk University, Brno (Czech Republic)

[2]TU München (Germany)

[3]University of Oxford (UK)

FSTTCS (Kanpur, India), 15 December, 2009

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \xhookrightarrow{1/2} YX \qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad Y \xhookrightarrow{1/3} \varepsilon$$

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \xhookrightarrow{1/2} YX \qquad\qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad\qquad Y \xhookrightarrow{1/3} \varepsilon$$

A run: X

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \xhookrightarrow{1/2} YX \qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad Y \xhookrightarrow{1/3} \varepsilon$$

A run: $X \overset{1/2}{\Longrightarrow} YX$

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \overset{1/2}{\hookrightarrow} YX \qquad Y \overset{2/3}{\hookrightarrow} X$$

$$X \overset{1/2}{\hookrightarrow} \varepsilon \qquad Y \overset{1/3}{\hookrightarrow} \varepsilon$$

A run: $X \overset{1/2}{\Longrightarrow} YX \overset{2/3}{\Longrightarrow} XX$

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \xhookrightarrow{1/2} YX \qquad\qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad\qquad Y \xhookrightarrow{1/3} \varepsilon$$

A run: $X \stackrel{1/2}{\Longrightarrow} YX \stackrel{2/3}{\Longrightarrow} XX \stackrel{1/2}{\Longrightarrow} YXX$

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \overset{1/2}{\hookrightarrow} YX \qquad\qquad Y \overset{2/3}{\hookrightarrow} X$$

$$X \overset{1/2}{\hookrightarrow} \varepsilon \qquad\qquad Y \overset{1/3}{\hookrightarrow} \varepsilon$$

A run: $X \overset{1/2}{\Longrightarrow} YX \overset{2/3}{\Longrightarrow} XX \overset{1/2}{\Longrightarrow} YXX \overset{1/3}{\Longrightarrow} XX$

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \xhookrightarrow{1/2} YX \qquad\qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad\qquad Y \xhookrightarrow{1/3} \varepsilon$$

A run: $X \xRightarrow{1/2} YX \xRightarrow{2/3} XX \xRightarrow{1/2} YXX \xRightarrow{1/3} XX \xRightarrow{1/2} X$

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \xhookrightarrow{1/2} YX \qquad\qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad\qquad Y \xhookrightarrow{1/3} \varepsilon$$

A run: $X \stackrel{1/2}{\Longrightarrow} YX \stackrel{2/3}{\Longrightarrow} XX \stackrel{1/2}{\Longrightarrow} YXX \stackrel{1/3}{\Longrightarrow} XX \stackrel{1/2}{\Longrightarrow} X \stackrel{1/2}{\Longrightarrow} \varepsilon$

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \overset{1/2}{\hookrightarrow} YX \qquad\qquad Y \overset{2/3}{\hookrightarrow} X$$

$$X \overset{1/2}{\hookrightarrow} \varepsilon \qquad\qquad Y \overset{1/3}{\hookrightarrow} \varepsilon$$

A run: $X \overset{1/2}{\Longrightarrow} YX \overset{2/3}{\Longrightarrow} XX \overset{1/2}{\Longrightarrow} YXX \overset{1/3}{\Longrightarrow} XX \overset{1/2}{\Longrightarrow} X \overset{1/2}{\Longrightarrow} \varepsilon$

(no control states in this talk)

Probabilistic Pushdown Systems

Probabilistic pushdown systems are a model for probabilistic procedural programs.

$$X \xhookrightarrow{1/2} YX \qquad\qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad\qquad Y \xhookrightarrow{1/3} \varepsilon$$

A run: $X \xRightarrow{1/2} YX \xRightarrow{2/3} XX \xRightarrow{1/2} YXX \xRightarrow{1/3} XX \xRightarrow{1/2} X \xRightarrow{1/2} \varepsilon$

(no control states in this talk)

Quantitative Properties of a run:

- its probability (product of the probabilities on the arrows)
- its time (number of steps till ε)
- its memory (longest configuration, “maximal stack height”)

Motivation

Probabilistic Pushdown Systems are a model for
probabilistic programs with unrestricted recursion.

Well-studied:

- model-checking for temporal logics [Etessami, Yannakakis, Esparza, Kučera, Mayr]
- long-run behaviour [Brázdil, Esparza, Kučera]
- games [Etessami, Yannakakis]

The basic quantities time and memory are random variables.

This talk is about the memory.

Motivation

Probabilistic Pushdown Systems are a model for
probabilistic programs with unrestricted recursion.

Well-studied:

- model-checking for temporal logics
  [Etessami, Yannakakis, Esparza, Kučera, Mayr]
- long-run behaviour [Brázdil, Esparza, Kučera]
- games [Etessami, Yannakakis]

The basic quantities time and memory are random variables.

This talk is about the memory.

What is the **distribution** of the memory?

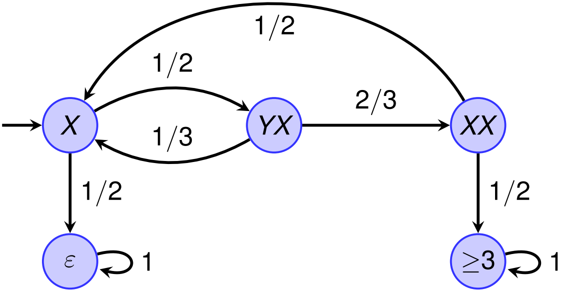
Computing the Distribution of the Memory

What is the probability of reaching height ≥ 3?

$$X \xhookrightarrow{1/2} YX \qquad\qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad\qquad Y \xhookrightarrow{1/3} \varepsilon$$

Idea: Construct the Markov chain with all stacks of height ≤ 2.



Computing the Distribution of the Memory

What is the probability of reaching height ≥ 3?

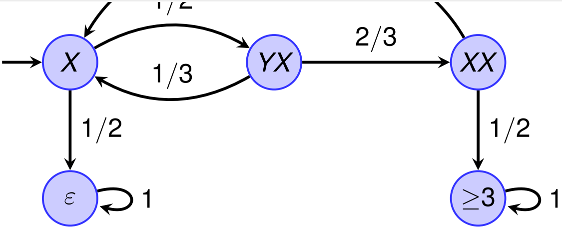
$$X \overset{1/2}{\hookrightarrow} YX \qquad\qquad Y \overset{2/3}{\hookrightarrow} X$$

$$X \overset{1/2}{\hookrightarrow} \varepsilon \qquad\qquad Y \overset{1/3}{\hookrightarrow} \varepsilon$$

Idea: Construct the Markov chain with all stacks of height ≤ 2.

Problem:

There may be 2^k states with height $\leq k$.



What is the probability of reaching height ≥ 3?

$$X \xhookrightarrow{1/2} YX \qquad\qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad\qquad Y \xhookrightarrow{1/3} \varepsilon$$

Idea: Construct the Markov chain with all stacks of height ≤ 2.

Problem:

There may be 2^k states with height $\leq k$.

Idea:

The Markov chain has a regular structure.
Exploit that.

Linear Equation Systems for the Distribution

$$X \xhookrightarrow{1/2} YX \qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad Y \xhookrightarrow{1/3} \varepsilon$$

Let $\boldsymbol{p}[k]_X :=$ probability of reaching height $\geq k$ if starting with X.
Compute $\boldsymbol{p}[k]_X$ by solving linear equations. For instance:

$$\boldsymbol{p}[10]_X = 1/2 \cdot (\boldsymbol{p}[9]_Y + \qquad \boldsymbol{p}[10]_X) + 1/2 \cdot 0$$

$$\boldsymbol{p}[10]_Y = 2/3 \cdot \boldsymbol{p}[10]_X + 1/3 \cdot 0$$

Linear Equation Systems for the Distribution

$$X \overset{1/2}{\hookrightarrow} YX \qquad Y \overset{2/3}{\hookrightarrow} X$$
$$X \overset{1/2}{\hookrightarrow} \varepsilon \qquad Y \overset{1/3}{\hookrightarrow} \varepsilon$$

Let $\boldsymbol{p}[k]_X :=$ probability of reaching height $\geq k$ if starting with X.
Compute $\boldsymbol{p}[k]_X$ by solving linear equations. For instance:

$$\boldsymbol{p}[10]_X = 1/2 \cdot (\boldsymbol{p}[9]_Y + \boldsymbol{t}[9]_Y \cdot \boldsymbol{p}[10]_X) + 1/2 \cdot 0$$
$$\boldsymbol{p}[10]_Y = 2/3 \cdot \boldsymbol{p}[10]_X + 1/3 \cdot 0$$

where $\boldsymbol{t}[k]_Y$ is the probability of terminating and having height $< k$.

Linear Equation Systems for the Distribution

$$X \xhookrightarrow{1/2} YX \qquad Y \xhookrightarrow{2/3} X$$

$$X \xhookrightarrow{1/2} \varepsilon \qquad Y \xhookrightarrow{1/3} \varepsilon$$

Let $\boldsymbol{p}[k]_X :=$ probability of reaching height $\geq k$ if starting with X.
Compute $\boldsymbol{p}[k]_X$ by solving linear equations. For instance:

$$\boldsymbol{p}[10]_X = 1/2 \cdot (\boldsymbol{p}[9]_Y + \boldsymbol{t}[9]_Y \cdot \boldsymbol{p}[10]_X) + 1/2 \cdot 0$$
$$\boldsymbol{p}[10]_Y = 2/3 \cdot \boldsymbol{p}[10]_X + 1/3 \cdot 0$$

where $\boldsymbol{t}[k]_Y$ is the probability of terminating and having height $< k$.

$$\boldsymbol{t}[10]_X = 1/2 \cdot \boldsymbol{t}[9]_Y \cdot \boldsymbol{t}[10]_X + 1/2 \cdot 1$$
$$\boldsymbol{t}[10]_Y = 2/3 \cdot \boldsymbol{t}[10]_X + 1/3 \cdot 1$$

Linear Equation Systems for the Distribution

Proposition

The vector $\mathbf{p}[k]$ can be computed
by setting up and solving linear equation systems.
It can be done in $\mathcal{O}(k \cdot |\Gamma|^3)$ arithmetic operations.

Linear Equation Systems for the Distribution

Proposition

The vector $\boldsymbol{p}[k]$ can be computed by setting up and solving linear equation systems. It can be done in $\mathcal{O}(k \cdot |\Gamma|^3)$ arithmetic operations.

Is this fast enough?

Linear Equation Systems for the Distribution

Proposition

The vector $\mathbf{p}[k]$ can be computed by setting up and solving linear equation systems.
It can be done in $\mathcal{O}(k \cdot |\Gamma|^3)$ arithmetic operations.

Is this fast enough? Depends.

Linear Equation Systems for the Distribution

Proposition

The vector $\mathbf{p}[k]$ can be computed
by setting up and solving linear equation systems.
It can be done in $\mathcal{O}(k \cdot |\Gamma|^3)$ arithmetic operations.

Is this fast enough? Depends.
What about the asymptotics for large k?

Linear Equation Systems for the Distribution

Proposition

The vector $\boldsymbol{p}[k]$ can be computed
by setting up and solving linear equation systems.
It can be done in $\mathcal{O}(k \cdot |\Gamma|^3)$ arithmetic operations.

Is this fast enough? Depends.
What about the asymptotics for large k?
What about the expectation of the memory consumption?

Linear Equation Systems for the Distribution

Proposition

The vector $\mathbf{p}[k]$ can be computed by setting up and solving linear equation systems. It can be done in $\mathcal{O}(k \cdot |\Gamma|^3)$ arithmetic operations.

Is this fast enough? Depends.
What about the asymptotics for large k?
What about the expectation of the memory consumption?

In the following we assume finite expectation of the memory (which is the most important case).

Putting the Equations in Matrix Form

After some (non-trivial but non-interesting) normalizations:

There is a (nonnegative) matrix $A(\boldsymbol{x}) \in \mathbb{R}^{\Gamma \times \Gamma}$
that depends on a (nonnegative) vector $\boldsymbol{x} \in \mathbb{R}^{\Gamma}$ such that:

- $A(\boldsymbol{x})$ increases monotonically with $\boldsymbol{x}$,
- $\boldsymbol{p}[k+1] = A(\boldsymbol{t}[k]) \cdot \boldsymbol{p}[k]$

Putting the Equations in Matrix Form

After some (non-trivial but non-interesting) normalizations:

There is a (nonnegative) matrix $A(\boldsymbol{x}) \in \mathbb{R}^{\Gamma \times \Gamma}$
that depends on a (nonnegative) vector $\boldsymbol{x} \in \mathbb{R}^{\Gamma}$ such that:

- $A(\boldsymbol{x})$ increases monotonically with $\boldsymbol{x}$,
- $\boldsymbol{p}[k+1] = A(\boldsymbol{t}[k]) \cdot \boldsymbol{p}[k]$

The sequence $\boldsymbol{t}[1] \leq \boldsymbol{t}[2] \leq \ldots$ converges to a limit $\boldsymbol{t}$,
where $\boldsymbol{t}$ is the vector of termination probabilities.

By monotonicity and continuity, the sequence
$A(\boldsymbol{t}[1]) \leq A(\boldsymbol{t}[2]) \leq \ldots$ also converges to a limit $A(\boldsymbol{t}) = A$.

Putting the Equations in Matrix Form

After some (non-trivial but non-interesting) normalizations:

There is a (nonnegative) matrix $A(\boldsymbol{x}) \in \mathbb{R}^{\Gamma \times \Gamma}$
that depends on a (nonnegative) vector $\boldsymbol{x} \in \mathbb{R}^{\Gamma}$ such that:

- $A(\boldsymbol{x})$ increases monotonically with $\boldsymbol{x}$,
- $\boldsymbol{p}[k+1] = A(\boldsymbol{t}[k]) \cdot \boldsymbol{p}[k]$

The sequence $\boldsymbol{t}[1] \leq \boldsymbol{t}[2] \leq \ldots$ converges to a limit $\boldsymbol{t}$,
where $\boldsymbol{t}$ is the vector of termination probabilities.

By monotonicity and continuity, the sequence
$A(\boldsymbol{t}[1]) \leq A(\boldsymbol{t}[2]) \leq \ldots$ also converges to a limit $A(\boldsymbol{t}) = A$.

It follows:

$$\boldsymbol{p}[k+1] = A(\boldsymbol{t}[k]) \cdot \boldsymbol{p}[k] \leq A \cdot \boldsymbol{p}[k] \leq A^k \cdot \boldsymbol{p}[1] = A^k \cdot \mathbf{1}$$

Approximating the Distribution

On the previous slide: $\boldsymbol{p}[k] \leq A^{k-1} \cdot \mathbf{1}$,
so $A^{k-1} \cdot \mathbf{1}$ is an upper bound on $\boldsymbol{p}[k]$.

Advantage: compute A^{k-1} by repeated squaring: $A, A^2, A^4, \ldots$
$\Longrightarrow$ only $\mathcal{O}(\log k \cdot |\Gamma|^3)$ operations
$\Longrightarrow$ safe overapproximation for large k

How tight is the overapproximation?

Approximating the Distribution

On the previous slide: $\boldsymbol{p}[k] \leq A^{k-1} \cdot \mathbf{1}$,
so $A^{k-1} \cdot \mathbf{1}$ is an upper bound on $\boldsymbol{p}[k]$.

Advantage: compute A^{k-1} by repeated squaring: $A, A^2, A^4, \ldots$
$\Longrightarrow$ only $\mathcal{O}(\log k \cdot |\Gamma|^3)$ operations
$\Longrightarrow$ safe overapproximation for large k

How tight is the overapproximation?

Proposition

There is a real number d with $0 < d \leq 1$ such that for all k:

$$d \cdot A^{k-1} \cdot \mathbf{1} \leq \boldsymbol{p}[k] \leq A^{k-1} \cdot \mathbf{1}\,.$$

Approximating the Distribution

On the previous slide:

Proposition

There is a real number d with $0 < d \leq 1$ such that for all k:

$$d \cdot A^{k-1} \cdot \mathbf{1} \leq \boldsymbol{p}[k] \leq A^{k-1} \cdot \mathbf{1}\,.$$

So, $\boldsymbol{p}[k]$ essentially depends on the spectral radius ρ of A.

Approximating the Distribution

On the previous slide:

Proposition

There is a real number d with $0 < d \leq 1$ such that for all k:

$$d \cdot A^{k-1} \cdot \mathbf{1} \leq \boldsymbol{p}[k] \leq A^{k-1} \cdot \mathbf{1}\,.$$

So, $\boldsymbol{p}[k]$ essentially depends on the spectral radius ρ of A.

Theorem

We have $\rho < 1$. The proposition above implies:

$$\boldsymbol{p}[k] \in \Theta(\rho^k)$$

Expected Memory

What is the **expectation** of the memory?

This expectation *EM* equals $\sum_{k=1}^{\infty} \boldsymbol{p}[k]_X$.

Expected Memory

What is the **expectation** of the memory?

This expectation *EM* equals $\sum_{k=1}^{\infty} \boldsymbol{p}[k]_X$.

EM can be (under-)approximated by $UM[\ell] := \sum_{k=1}^{\ell} \boldsymbol{p}[k]_X$.

Theorem

The sequence $(UM[\ell])_\ell$ *converges to EM.*
More precisely, one can compute $a > 0$ *and* $0 < b < 1$ *with*

$$EM - UM[\ell] \leq a \cdot b^{\ell} \ ,$$

so the error decays exponentially.

Finiteness of the Expectation

Most presented results hold if the expectation is finite.
How to decide whether the expectation is finite?

- $X \xhookrightarrow{2/3} XX,\ X \xhookrightarrow{1/3} \varepsilon$: $\quad \mathcal{P}(M = \infty) = 1/2 > 0$, so $EM = \infty$
- $X \xhookrightarrow{1/2} XX,\ X \xhookrightarrow{1/2} \varepsilon$:

Finiteness of the Expectation

Most presented results hold if the expectation is finite.
How to decide whether the expectation is finite?

- $X \xhookrightarrow{2/3} XX, \; X \xhookrightarrow{1/3} \varepsilon$: $\quad \mathcal{P}(M = \infty) = 1/2 > 0$, so $EM = \infty$
- $X \xhookrightarrow{1/2} XX, \; X \xhookrightarrow{1/2} \varepsilon$: $\quad \mathcal{P}(M = \infty) = 0$, but still $EM = \infty$

Finiteness of the Expectation

Most presented results hold if the expectation is finite.
How to decide whether the expectation is finite?

- $X \xhookrightarrow{2/3} XX, \; X \xhookrightarrow{1/3} \varepsilon$: $\quad \mathcal{P}(M = \infty) = 1/2 > 0$, so $EM = \infty$
- $X \xhookrightarrow{1/2} XX, \; X \xhookrightarrow{1/2} \varepsilon$: $\quad \mathcal{P}(M = \infty) = 0$, but still $EM = \infty$
- $X \xhookrightarrow{1/3} XX, \; X \xhookrightarrow{2/3} \varepsilon$: $\quad \mathcal{P}(M = \infty) = 0$, and $EM \approx 1.61$

Finiteness of the Expectation

Most presented results hold if the expectation is finite.
How to decide whether the expectation is finite?

- $X \xhookrightarrow{2/3} XX,\ X \xhookrightarrow{1/3} \varepsilon$: $\quad \mathcal{P}(M = \infty) = 1/2 > 0$, so $EM = \infty$
- $X \xhookrightarrow{1/2} XX,\ X \xhookrightarrow{1/2} \varepsilon$: $\quad \mathcal{P}(M = \infty) = 0$, but still $EM = \infty$
- $X \xhookrightarrow{1/3} XX,\ X \xhookrightarrow{2/3} \varepsilon$: $\quad \mathcal{P}(M = \infty) = 0$, and $EM \approx 1.61$

Theorem

Whether EM is finite can be decided

- *in polynomial time for pushdown systems without states*
- *in polynomial space for general pushdown systems.*

The problem is PosSLP-hard (therefore unlikely in P)
for general pushdown systems.

Conclusions

- Probabilistic pushdown systems model
  recursive probabilistic programs.
- We studied one basic random variable, the memory.
- Its distribution can be computed
  - naively (using an exponential-sized Markov chain),
  - by solving linear equation systems,
  - by efficiently computing overapproximations of the memory.
- The expectation can be approximated,
  and the error decays exponentially.
- Whether the expectation is finite can be decided in
  polynomial time for pushdown systems without states.
- Open question: How to decide whether
  the expectation exceeds a given bound?

End of Talk

Thank you!