Tail Probabilities for Randomized Program Runtimes via Martingales for Higher Moments

Satoshi Kura[1,2] Natsuki Urabe[1]
Ichiro Hasuo[1,2]

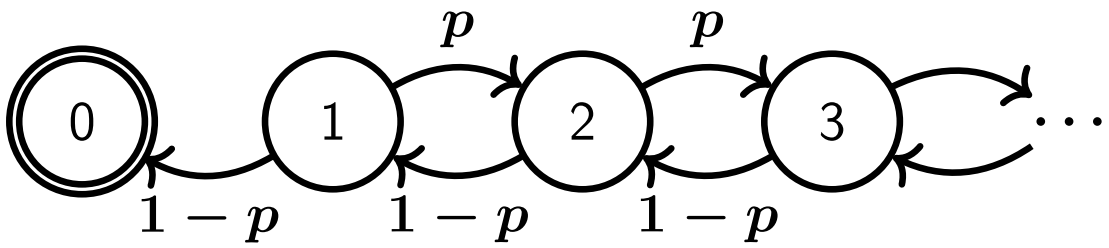
[1]National Institute of Informatics, Tokyo, Japan

[2]The Graduate University for Advanced Studies (SOKENDAI), Kanagawa, Japan

April 10, 2019

Our question

"What is an upper bound of the tail probability?"

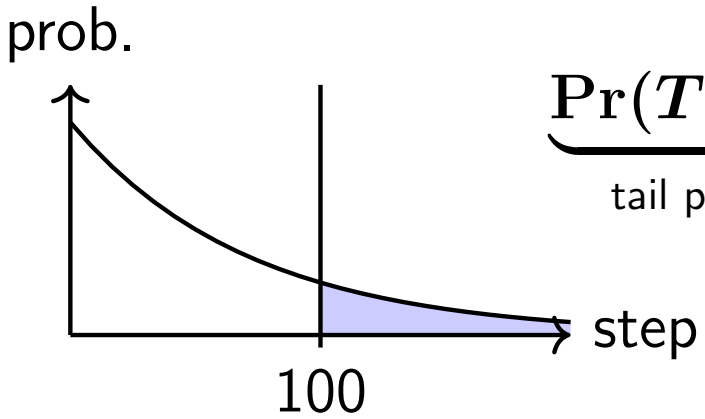


How likely is it to terminate within 100 steps?

(e.g. at least 90%)

How unlikely is it to not terminate within 100 steps?

(e.g. at most 10%)



$$\underbrace{\Pr(T \geq 100)}_{\text{tail probability}} \leq ??$$

Related work

Supermartingale-based approach

- Proving almost-sure termination

  [Chakarov & Sankaranarayanan, CAV'13]

- Overapproximating tail probabilities:

$$\mathbf{Pr}(T \geq d) \leq ??$$

  [Chatterjee & Fu, arxiv preprint], [Chatterjee et al., TOPLAS'18]

  - Azuma's, Hoeffding's and Bernstein's inequalities
  - Markov's inequality (wider applicability)

$$\mathbf{Pr}(T \geq d) \leq \frac{\mathbf{E}[T]}{d}$$

Our approach

- Aim: overapproximating tail probabilities:

$$\mathbf{Pr}(T \geq d) \leq ??$$

- Corollary of Markov's inequality

$$\mathbf{Pr}(T \geq d) \leq \frac{\mathbf{E}[T^k]}{d^k}$$

- Extends ranking supermartingale for <u>higher moments</u> $\mathbf{E}[T^k]$ $(k = 1, 2, \dots)$

Our workflow

randomized program

$\downarrow$

$\boxed{\text{our supermartingales}}$

$\downarrow$

upper bounds of higher moments

$$\left(\mathbf{E}[T], \ldots, \mathbf{E}[T^K]\right) \leq (u_1, \ldots, u_K)$$

$\downarrow$

deadline $d \longrightarrow \boxed{\text{concentration inequality}}$

$\downarrow$

upper bound of tail probability

$$\mathbf{Pr}(T \geq d) \leq ?$$

Our workflow

randomized program

↓

our supermartingales

↓

upper bounds of higher moments

$$\left(\mathbf{E}[T], \ldots, \mathbf{E}[T^K]\right) \leq (u_1, \ldots, u_K)$$

↓

deadline d → concentration inequality

↓

upper bound of tail probability

$$\mathbf{Pr}(T \geq d) \leq ?$$

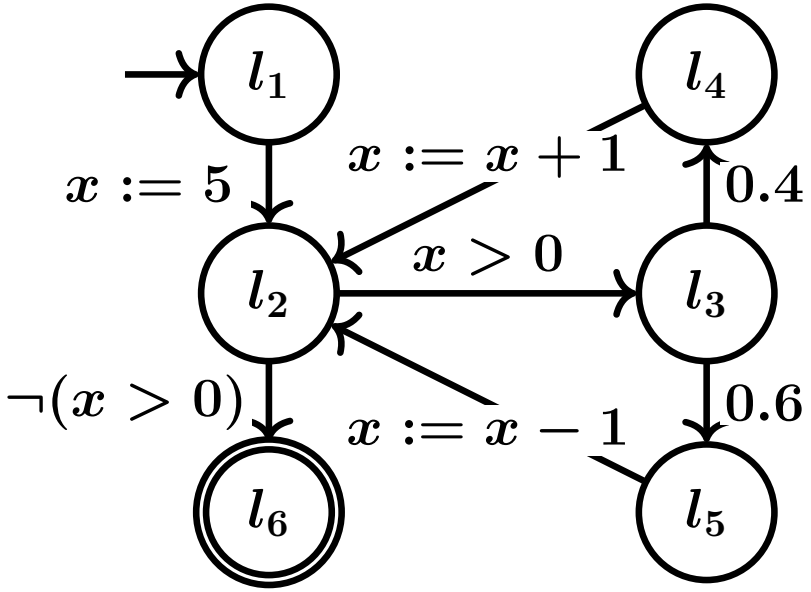
Randomized program

✓ sampling

✓ (demonic/termination avoiding) nondeterminism

Given as a pCFG (probabilistic control flow graph).

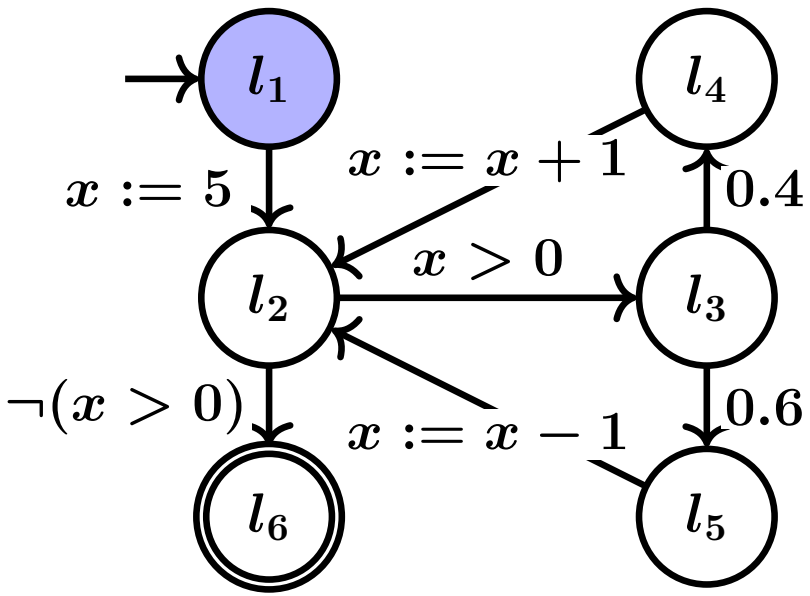
```
x := 5;
while x > 0 do
  if prob(0.4) then
    x := x + 1
  else
    x := x - 1
  fi
od
```



Semantics

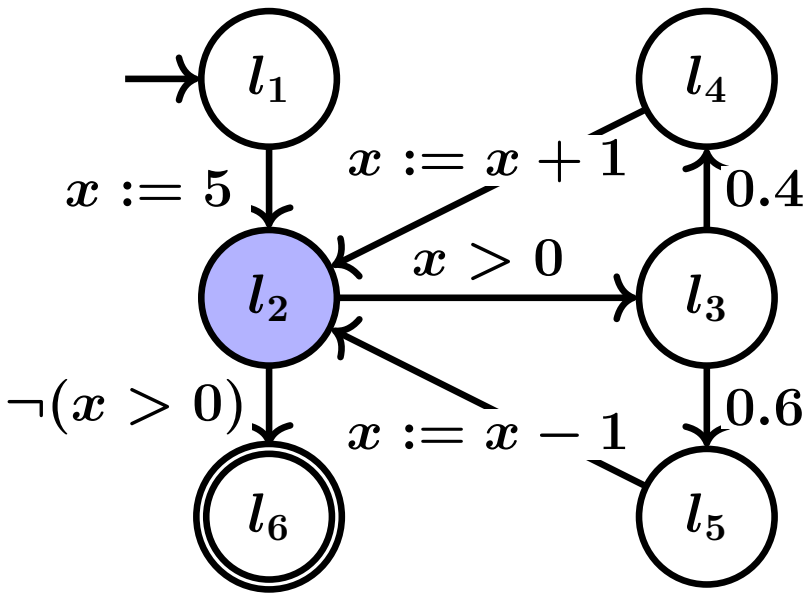
- Configuration: $(l, \vec{x}) \in L \times \mathbb{R}^V$
  - L: finite set of locations
  - V: finite set of program variables
- Run: sequence of configurations

$$\downarrow$$
$$(l_1, [x \mapsto 0])$$



Semantics

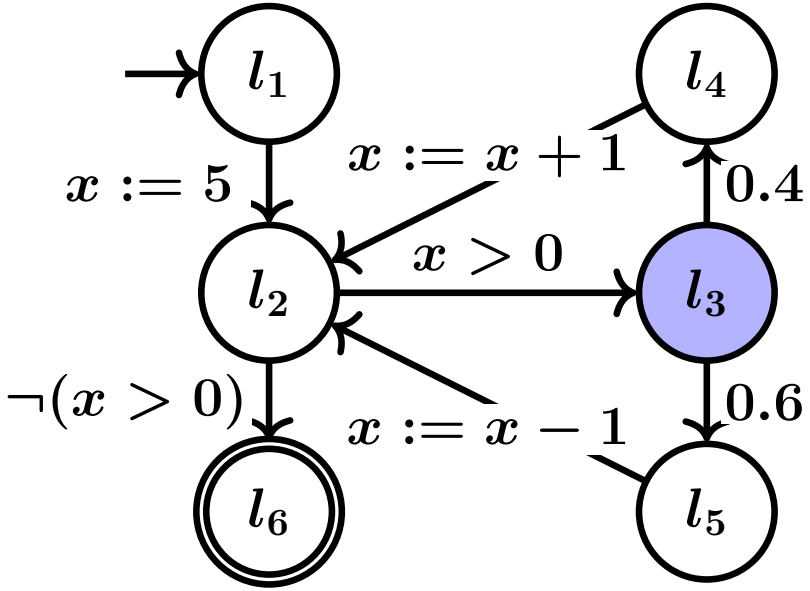
- Configuration: $(l, \vec{x}) \in L \times \mathbb{R}^V$
  - L: finite set of locations
  - V: finite set of program variables
- Run: sequence of configurations



$$\begin{array}{c} \downarrow \\ (l_1, [x \mapsto 0]) \\ 1 \downarrow \\ (l_2, [x \mapsto 5]) \end{array}$$

Semantics

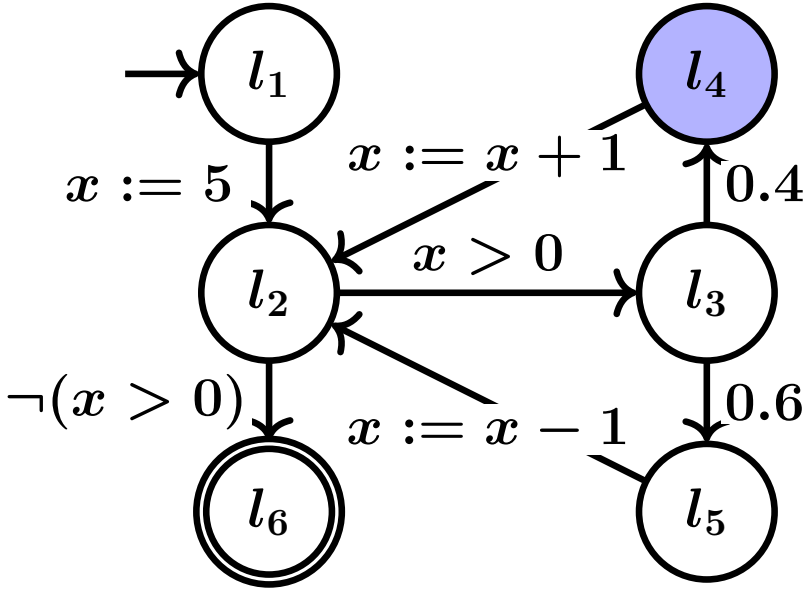
- Configuration: $(l, \vec{x}) \in L \times \mathbb{R}^V$
  - L: finite set of locations
  - V: finite set of program variables
- Run: sequence of configurations

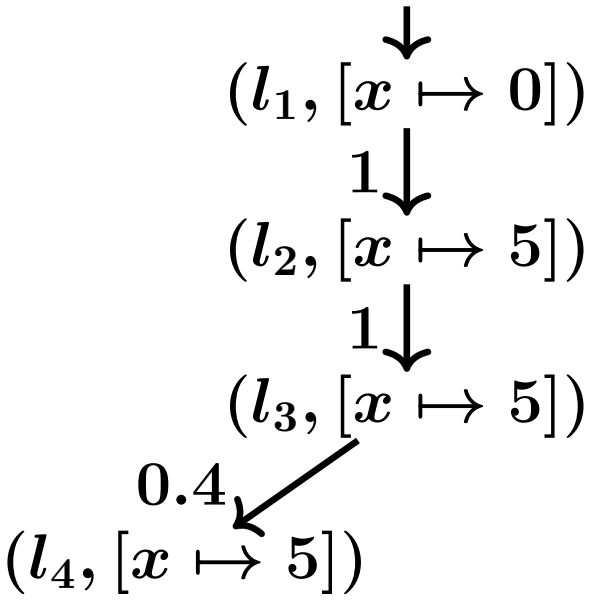


$$\begin{array}{c} \downarrow \\ (l_1, [x \mapsto 0]) \\ 1 \downarrow \\ (l_2, [x \mapsto 5]) \\ 1 \downarrow \\ (l_3, [x \mapsto 5]) \end{array}$$

Semantics

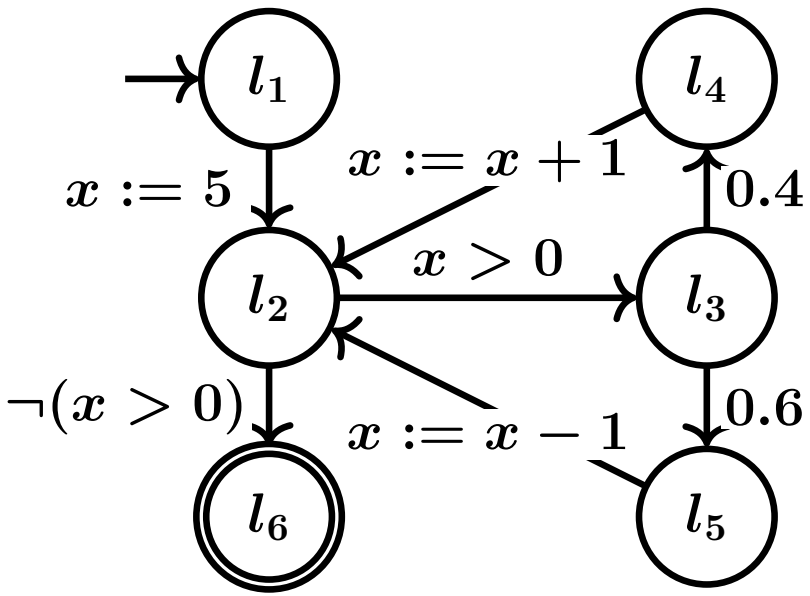
- Configuration: $(l, \vec{x}) \in L \times \mathbb{R}^V$
  - L: finite set of locations
  - V: finite set of program variables
- Run: sequence of configurations





Semantics

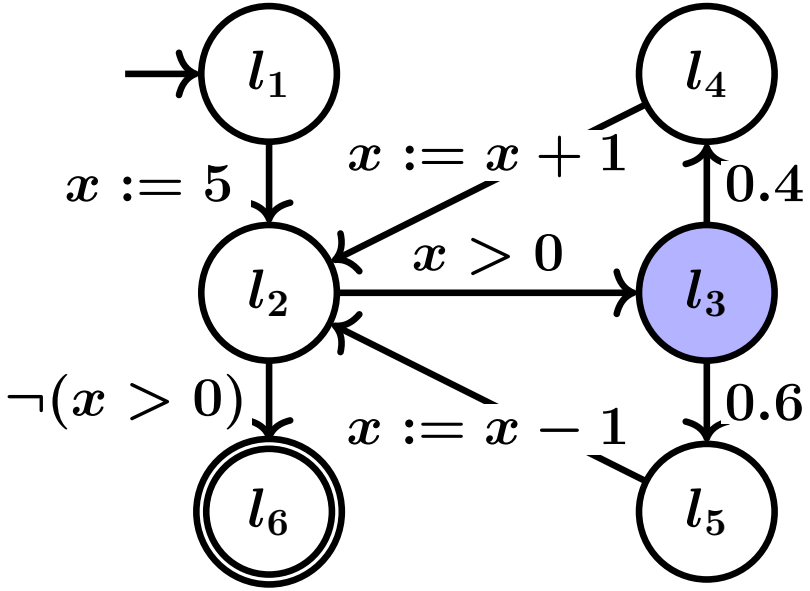
- Configuration: $(l, \vec{x}) \in L \times \mathbb{R}^V$
  - L: finite set of locations
  - V: finite set of program variables
- Run: sequence of configurations



$$
\begin{array}{c}
\downarrow \\
(l_1, [x \mapsto 0]) \\
1 \downarrow \\
(l_2, [x \mapsto 5]) \\
1 \downarrow \\
(l_3, [x \mapsto 5]) \\
0.4 \swarrow \\
(l_4, [x \mapsto 5]) \\
\downarrow \\
\vdots
\end{array}
$$

Semantics

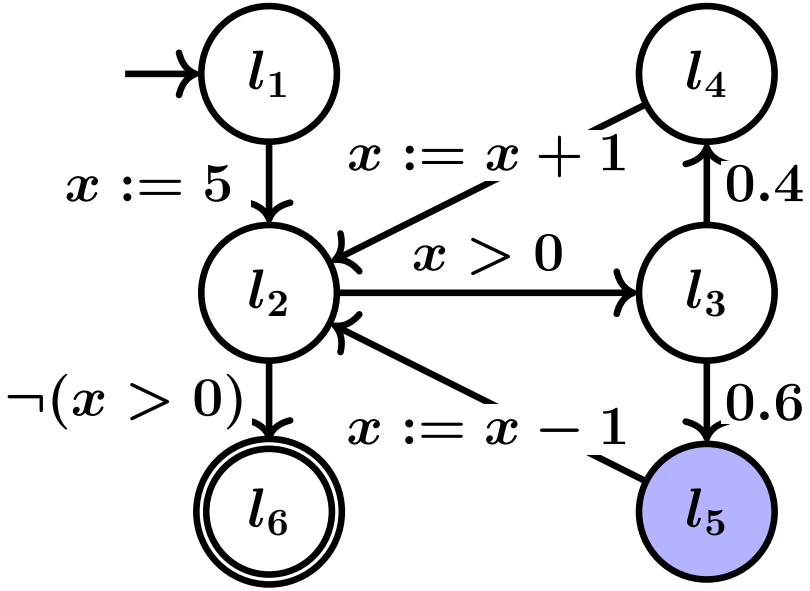
- Configuration: $(l, \vec{x}) \in L \times \mathbb{R}^V$
  - L: finite set of locations
  - V: finite set of program variables
- Run: sequence of configurations

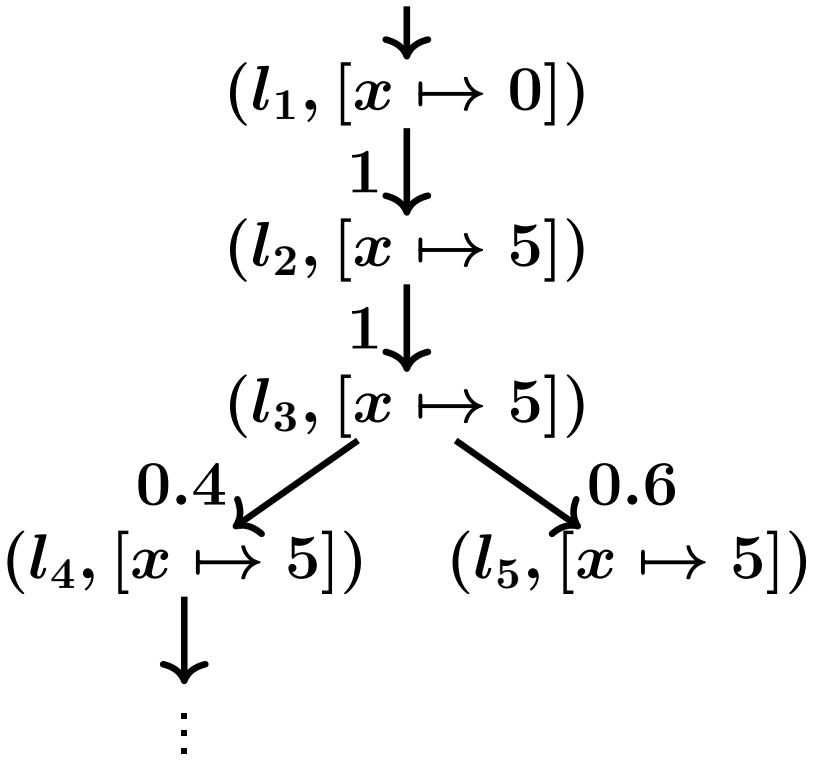


$$
\begin{array}{c}
\downarrow \\
(l_1, [x \mapsto 0]) \\
1 \downarrow \\
(l_2, [x \mapsto 5]) \\
1 \downarrow \\
(l_3, [x \mapsto 5]) \\
0.4 \swarrow \\
(l_4, [x \mapsto 5]) \\
\downarrow \\
\vdots
\end{array}
$$

Semantics

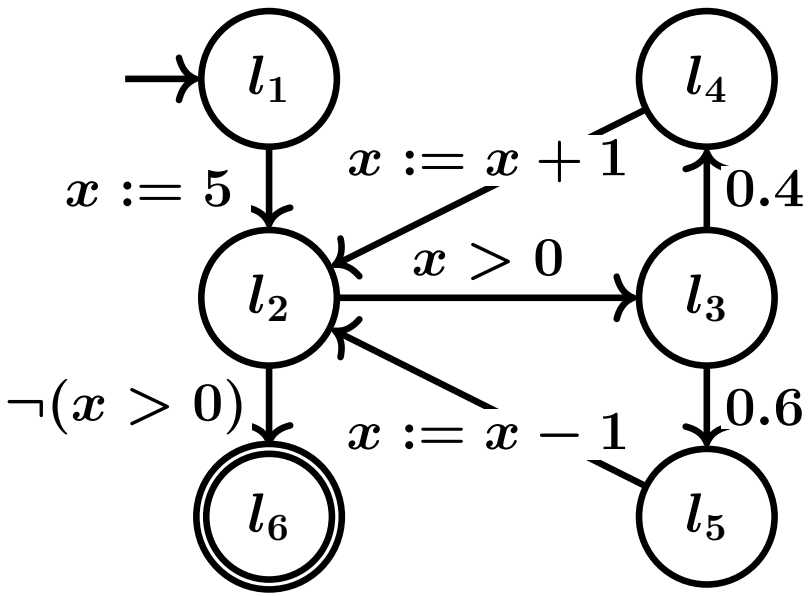
- Configuration: $(l, \vec{x}) \in L \times \mathbb{R}^V$
  - L: finite set of locations
  - V: finite set of program variables
- Run: sequence of configurations

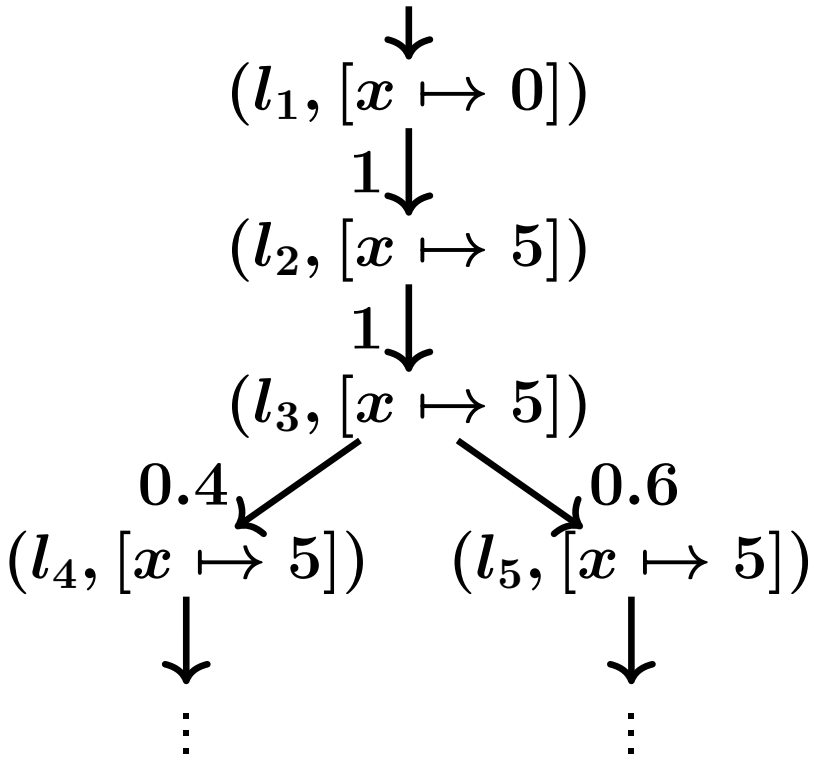




Semantics

- Configuration: $(l, \vec{x}) \in L \times \mathbb{R}^V$
  - L: finite set of locations
  - V: finite set of program variables
- Run: sequence of configurations





Our workflow

randomized program

↓

$\boxed{\text{our supermartingales}}$

↓

upper bounds of higher moments

$$\left(\mathbf{E}[T], \ldots, \mathbf{E}[T^K]\right) \leq (u_1, \ldots, u_K)$$

↓

deadline d → $\boxed{\text{concentration inequality}}$

↓

upper bound of tail probability

$$\mathbf{Pr}(T \geq d) \leq ?$$

Ranking function

[Floyd, '67]

$$r : L \times \mathbb{R}^V \to \mathbb{N} \cup \{\infty\}$$

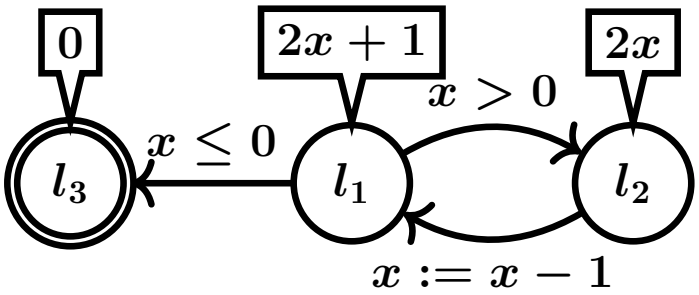
For each transition, r decreases by (at least) 1:

$$(l, \vec{x}) \mapsto (l', \vec{x'}) \implies r(l', \vec{x'}) \leq r(l, \vec{x}) - 1$$

Theorem

If $r(l, \vec{x}) < \infty$, then the program is terminating from $(l, \vec{x})$ within $r(l, \vec{x})$ steps.

```
x := 5;
while x > 0 do
  x := x - 1
od
```



Ranking supermartingale

[Chakarov & Sankaranarayanan, CAV'13]

$$\eta : L \times \mathbb{R}^V \to [0, \infty]$$

For each transition, η decreases by (at least) 1 "on average":

$$(\mathbb{X}\eta)(l, \vec{x}) \leq \eta(l, \vec{x}) - 1 \quad \text{for each } (l, \vec{x})$$

where $\mathbb{X}$ is next-time operator (the expected value after one transition):

$$(\mathbb{X}\eta)(l, \vec{x}) := E[\eta(l', \vec{x}') \mid (l, \vec{x}) \mapsto (l', \vec{x}')].$$

Ranking supermartingale

Theorem

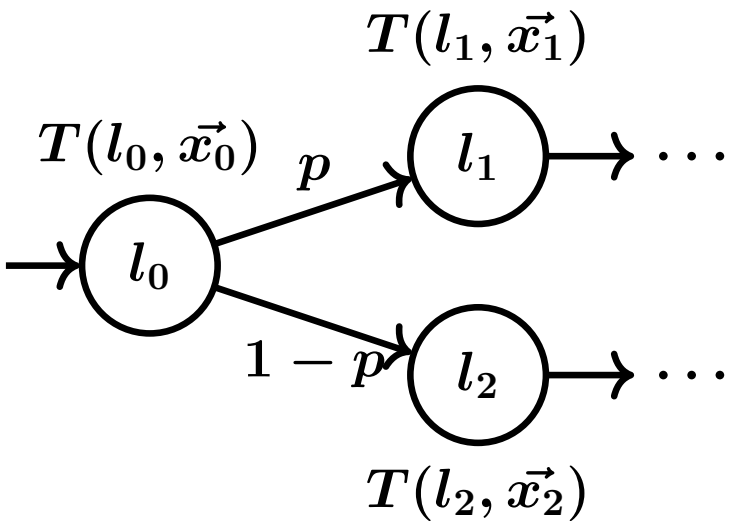
If $\eta(l, \vec{x}) < \infty$, then the program is (positively) almost surely terminating from $(l, \vec{x})$ with the expected runtime $\leq \eta(l, \vec{x})$ steps.

This can be explained lattice-theoretically.

- Expected runtime is a lfp
- Ranking supermartingale is a prefixed point

Runtime before and after transition

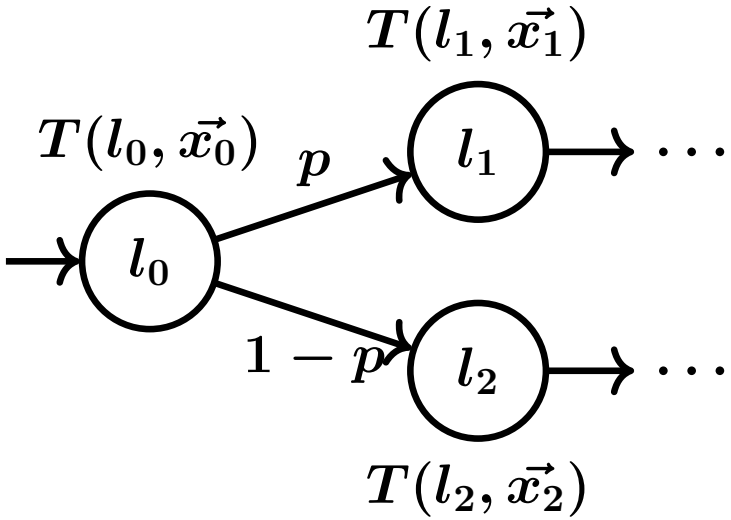
Let $T(l, \vec{x})$ be a random variable representing the runtime from $(l, \vec{x})$.



Runtime from $(l_0, \vec{x_0})$:

- $T(l_1, \vec{x_1}) + 1$ with probability p
- $T(l_2, \vec{x_2}) + 1$ with probability $1 - p$

Expected runtime is a fixed point



$$
\begin{aligned}
&\mathbf{E}[T](l_0, \vec{x_0}) \\
&= p\mathbf{E}[T(l_1, \vec{x_1}) + 1] + (1-p)\mathbf{E}[T(l_2, \vec{x_2}) + 1] \\
&= p(\mathbf{E}[T(l_1, \vec{x_1})] + 1) + (1-p)(\mathbf{E}[T(l_2, \vec{x_2})] + 1) \\
&= \mathbf{E}\left[\mathbf{E}[T(l', \vec{x'})] + 1 \mid (l_0, \vec{x_0}) \mapsto (l', \vec{x'})\right] \\
&= (\mathbb{X}(\mathbf{E}[T] + 1))(l_0, \vec{x_0})
\end{aligned}
$$

where $\mathbf{E}[T] := \lambda(l, \vec{x}).\ \mathbf{E}[T(l, \vec{x})]$.

Expected runtime is lfp

$$\mathbf{E}[\boldsymbol{T}] = \mathbb{X}(\mathbf{E}[\boldsymbol{T}] + \mathbf{1})$$

In fact, $\mathbf{E}[\boldsymbol{T}]$ is the "least" fixed point of

$$\boldsymbol{F}_1(\boldsymbol{\eta}) := \mathbb{X}(\boldsymbol{\eta} + \mathbf{1}).$$

- $\boldsymbol{F}_1$ is a monotone function on the complete lattice $[\mathbf{0}, \boldsymbol{\infty}]^{\boldsymbol{L} \times \mathbb{R}^{\boldsymbol{V}}}$
- $\boldsymbol{F}_1$ adds 1 unit of time, and then calculate the expected value after one transition

Ranking supermartingale is prefixed point

η is a ranking supermartingale
$\iff$ η is a prefixed point of F_1

$$F_1\eta = \mathbb{X}(\eta + 1) \leq \eta$$

Theorem (Knaster–Tarski)

Let L be a complete lattice and $F : L \to L$ be a monotone function. The least fixed point μF is the least prefixed point. Therefore we have

$$F\eta \leq \eta \implies \mu F \leq \eta.$$

It follows that

$$\eta \text{ is a ranking supermartingale} \implies \mathrm{E}[T] \leq \eta.$$

Our supermartingale

	[Chakarov & Sankaranarayanan, CAV'13]
lattice	$\boldsymbol{L} \times \mathbb{R}^{\boldsymbol{V}} \to [\mathbf{0}, \boldsymbol{\infty}]$
monotone function $\boldsymbol{F}$	$\boldsymbol{F_1}$
lfp $\boldsymbol{\mu F}$	$\mathbf{E}[\boldsymbol{T}]$
prefixed point $\boldsymbol{F\eta} \leq \boldsymbol{\eta}$	ranking supermartingale $\boldsymbol{\eta}$
Knaster–Tarski $\boldsymbol{\mu F} \leq \boldsymbol{\eta}$	$\mathbf{E}[\boldsymbol{T}] \leq \boldsymbol{\eta}$

† for a pCFG without nondeterminism

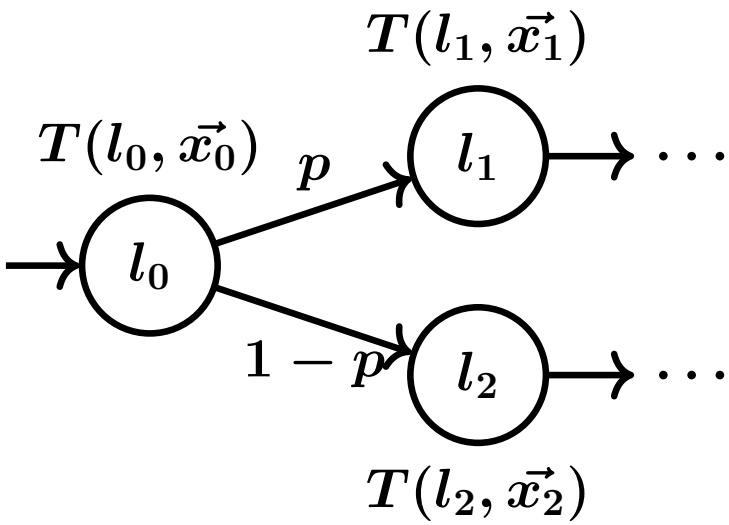
Our supermartingale

	[Chakarov & Sankaranarayanan, CAV'13]	Our supermartingale
lattice	$\boldsymbol{L} \times \mathbb{R}^{\boldsymbol{V}} \to [\mathbf{0}, \infty]$	$\boldsymbol{L} \times \mathbb{R}^{\boldsymbol{V}} \to [\mathbf{0}, \infty]^{\boldsymbol{K}}$
monotone function $\boldsymbol{F}$	$\boldsymbol{F_1}$	$\boldsymbol{F_K}$
lfp $\boldsymbol{\mu F}$	$\mathbf{E}[\boldsymbol{T}]$	$(\mathbf{E}[\boldsymbol{T}], \ldots, \mathbf{E}[\boldsymbol{T}^{\boldsymbol{K}}])^{\dagger}$
prefixed point $\boldsymbol{F\eta} \leq \boldsymbol{\eta}$	ranking supermartingale $\boldsymbol{\eta}$	ranking supermartingale for higher moments $\vec{\boldsymbol{\eta}}$
Knaster–Tarski $\boldsymbol{\mu F} \leq \boldsymbol{\eta}$	$\mathbf{E}[\boldsymbol{T}] \leq \boldsymbol{\eta}$	$(\mathbf{E}[\boldsymbol{T}], \ldots, \mathbf{E}[\boldsymbol{T}^{\boldsymbol{K}}]) \leq \vec{\boldsymbol{\eta}}$

†for a pCFG without nondeterminism

Runtime before and after transition

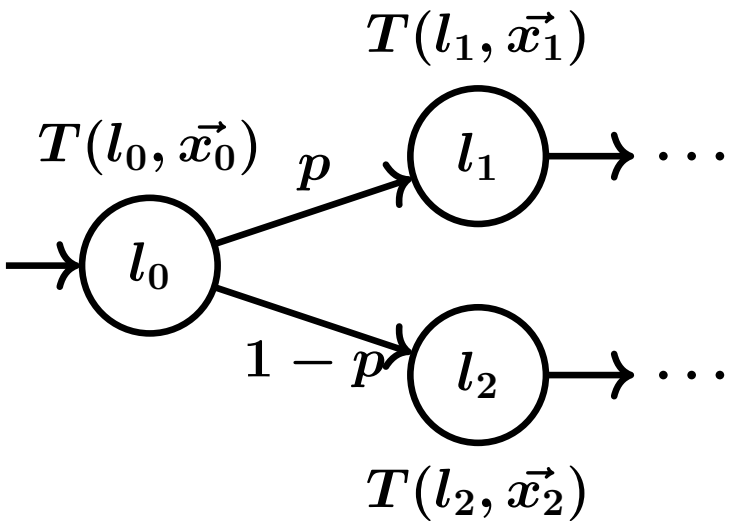
Let $T(l, \vec{x})$ be a random variable representing the runtime from $(l, \vec{x})$.



Runtime from $(l_0, \vec{x_0})$:

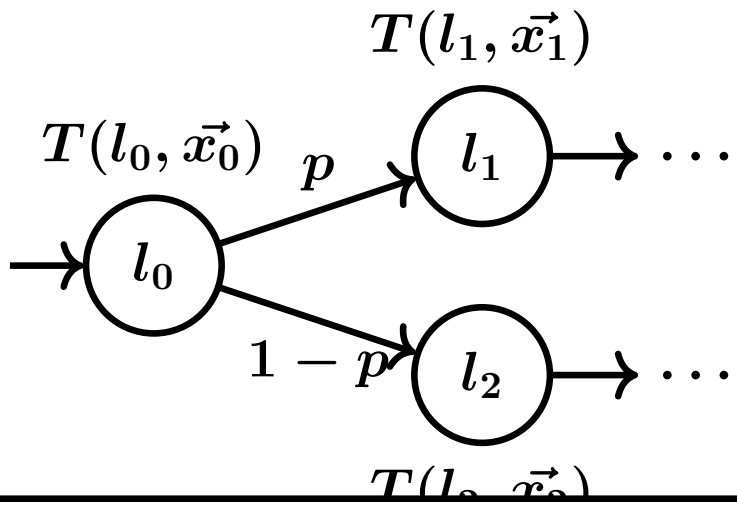
- $T(l_1, \vec{x_1}) + 1$ with probability p
- $T(l_2, \vec{x_2}) + 1$ with probability $1 - p$

Characterizing $\mathbf{E}[T^2]$ as lfp?



$$
\begin{aligned}
&\mathbf{E}[T^2](l_0, \vec{x_0}) \\
&= p\mathbf{E}[\left(T(l_1, \vec{x_1}) + 1\right)^2] \\
&\qquad + (1-p)\mathbf{E}[(T(l_2, \vec{x_2}) + 1)^2] \\
&= \left(\mathbb{X}(\mathbf{E}[T^2] + 2\underline{\underline{\mathbf{E}[T]}} + 1)\right)(l_0, \vec{x_0})
\end{aligned}
$$

Characterizing $\mathbf{E}[T^2]$ as lfp?



Calculate $\mathbf{E}[T]$ and $\mathbf{E}[T^2]$ simultaneously

$$= \left(\mathbb{X}(\mathbf{E}[T^2] + 2\underline{\underline{\mathbf{E}[T]}} + 1)\right)(l_0, \vec{x_0})$$

Characterizing $\mathbf{E}[T]$ and $\mathbf{E}[T^2]$ as lfp

$$\begin{pmatrix} \mathbf{E}[T] \\ \mathbf{E}[T^2] \end{pmatrix} = \mathbb{X}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{E}[T] \\ \mathbf{E}[T^2] \end{pmatrix} \right)$$

In fact, $(\mathbf{E}[T], \mathbf{E}[T^2])$ is the "least" fixed point of

$$F_2 \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} := \mathbb{X}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \right)$$

where

- $\eta_1, \eta_2 : L \times \mathbb{R}^V \to [0, \infty]$
- $\mathbf{E}[T] = \lambda(l, \vec{x}).\, \mathbf{E}[T(l, \vec{x})]$
- $\mathbf{E}[T^2] = \lambda(l, \vec{x}).\, \mathbf{E}[\big(T(l, \vec{x})\big)^2]$

Characterizing higher moments as lfp

In the same way, we can define

$$\boldsymbol{F_K} : (\boldsymbol{L} \times \mathbb{R}^{\boldsymbol{V}} \to [\mathbf{0}, \infty]^{\boldsymbol{K}}) \to (\boldsymbol{L} \times \mathbb{R}^{\boldsymbol{V}} \to [\mathbf{0}, \infty]^{\boldsymbol{K}})$$

that characterizes higher moments of runtime.

Lemma

- *For a pCFG without nondeterminism,*

$$(\mathbf{E}[\boldsymbol{T}], \ldots, \mathbf{E}[\boldsymbol{T}^{\boldsymbol{K}}]) = \boldsymbol{\mu F_K}.$$

- *In general (with nondeterminism),*

$$(\mathbf{E}[\boldsymbol{T}], \ldots, \mathbf{E}[\boldsymbol{T}^{\boldsymbol{K}}]) \leq \boldsymbol{\mu F_K}.$$

Supermartingale is a prefixed point

Definition

A **ranking supermartingale for K-th moment** is a prefixed point $\vec{\eta} = (\eta_1, \ldots, \eta_K)$ of F_K.

$$F_K \vec{\eta} \leq \vec{\eta}$$

By the Knaster–Tarski theorem, $\vec{\eta}$ gives an upper bound (even with nondeterminism).

$$\begin{pmatrix} \mathbf{E}[T] \\ \vdots \\ \mathbf{E}[T^K] \end{pmatrix} \leq \mu F_K \leq \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_K \end{pmatrix}$$

Our workflow

randomized program

↓

$\boxed{\text{our supermartingales}}$

↓

upper bounds of higher moments

$$\left(\mathbf{E}[T], \ldots, \mathbf{E}[T^K]\right) \leq (u_1, \ldots, u_K)$$

↓

deadline d → $\boxed{\text{concentration inequality}}$

↓

upper bound of tail probability

$$\mathbf{Pr}(T \geq d) \leq ?$$

Problem

Assume

- $\boldsymbol{d} > \mathbf{0}$,
- $\boldsymbol{T}$ is a nonnegative random variable,
- $\begin{pmatrix} \mathbf{E}[\boldsymbol{T}] \\ \vdots \\ \mathbf{E}[\boldsymbol{T}^{\boldsymbol{K}}] \end{pmatrix} \leq \begin{pmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_{\boldsymbol{K}} \end{pmatrix}$,
- but we do not know the exact values of $\mathbf{E}[\boldsymbol{T}], \ldots, \mathbf{E}[\boldsymbol{T}^{\boldsymbol{K}}]$.

How to obtain an upper bound of $\boldsymbol{P}(\boldsymbol{T} \geq \boldsymbol{d})$?

If $K = 1$ ...

Theorem (Markov's inequality)

If T is a nonnegative r.v. and $d > 0$,

$$\Pr(T \geq d) \leq \frac{\mathrm{E}[T]}{d}.$$

By $\mathrm{E}[T] \leq u_1$,

$$\Pr(T \geq d) \leq \frac{\mathrm{E}[T]}{d} \leq \frac{u_1}{d}.$$

General case

- For any $k \in \{1, \ldots, K\}$,

$$\begin{aligned}\Pr(T \geq d) &= \Pr(T^k \geq d^k) \\ &\leq \frac{\mathrm{E}[T^k]}{d^k} \\ &\leq \frac{u_k}{d^k}\end{aligned}$$

- ("0-th" moment)

$$\Pr(T \geq d) \leq 1 = \frac{\mathrm{E}[T^0]}{d^0}$$

Concentration inequality we used

$$\Pr(T \geq d) \leq \min_{k=0,\ldots,K} \frac{u_k}{d^k}$$

where

- $d > 0$
- T is a nonnegative random variable
- $\begin{pmatrix} \mathrm{E}[T] \\ \vdots \\ \mathrm{E}[T^K] \end{pmatrix} \leq \begin{pmatrix} u_1 \\ \vdots \\ u_K \end{pmatrix}$
- $u_0 = 1$

Moreover, this gives the "optimal" upper bound under the above conditions.

Our workflow

randomized program

$\downarrow$

$\boxed{\text{our supermartingales}}$

$\downarrow$

upper bounds of higher moments

$$\left(\mathbf{E}[T], \ldots, \mathbf{E}[T^K]\right) \leq (u_1, \ldots, u_K)$$

$\downarrow$

deadline $d \longrightarrow \boxed{\text{concentration inequality}}$

$\downarrow$

upper bound of tail probability

$$\mathbf{Pr}(T \geq d) \leq ?$$

Synthesis (linear template)

Based on [Chakarov & Sankaranarayanan, CAV'13]

- Input: a pCFG with initial config $(l_{\mathrm{init}}, \vec{x}_{\mathrm{init}})$
- Output: an upper bound of $\mathbf{E}[T^K](l_{\mathrm{init}}, \vec{x}_{\mathrm{init}})$

Assume that $\vec{\eta} = (\eta_1, \ldots, \eta_K)$ is linear:

$$\eta_k(l, \vec{x}) = \vec{a}_{k,l} \cdot \vec{x} + b_{k,l} \qquad (k = 1, \ldots, K)$$

Determine $\vec{a}_{k,l}$, $b_{k,l}$ by solving the LP problem:

- minimize: $\eta_K(l_{\mathrm{init}}, \vec{x}_{\mathrm{init}})$
- subject to: ranking supermartingale condition (using Farkas' lemma)

Then we have

$$\mathbf{E}[T^K](l_{\mathrm{init}}, \vec{x}_{\mathrm{init}}) \leq \min \eta_K(l_{\mathrm{init}}, \vec{x}_{\mathrm{init}})$$

Synthesis (polynomial template)

Based on [Chatterjee et al., CAV'16]

- Input: a pCFG with initial config $(l_{\text{init}}, \vec{x}_{\text{init}})$
- Output: an upper bound of $\mathbf{E}[T^K](l_{\text{init}}, \vec{x}_{\text{init}})$

Assume that $\vec{\eta} = (\eta_1, \ldots, \eta_K)$ is polynomial.
Determine coefficients by solving the SDP problem:

- minimize: $\eta_K(l_{\text{init}}, \vec{x}_{\text{init}})$
- subject to: ranking supermartingale condition (using Positivstellensatz)

Then we have

$$\mathbf{E}[T^K](l_{\text{init}}, \vec{x}_{\text{init}}) \leq \min \eta_K(l_{\text{init}}, \vec{x}_{\text{init}})$$

Experiments

- Implementation based on linear/polynomial templates
- Tested 7 example programs
  - 2 coupon collector's problems
  - 5 random walks (some of them include nondeterminism)
- (degree of polynomial template) $\leq \mathbf{3}$

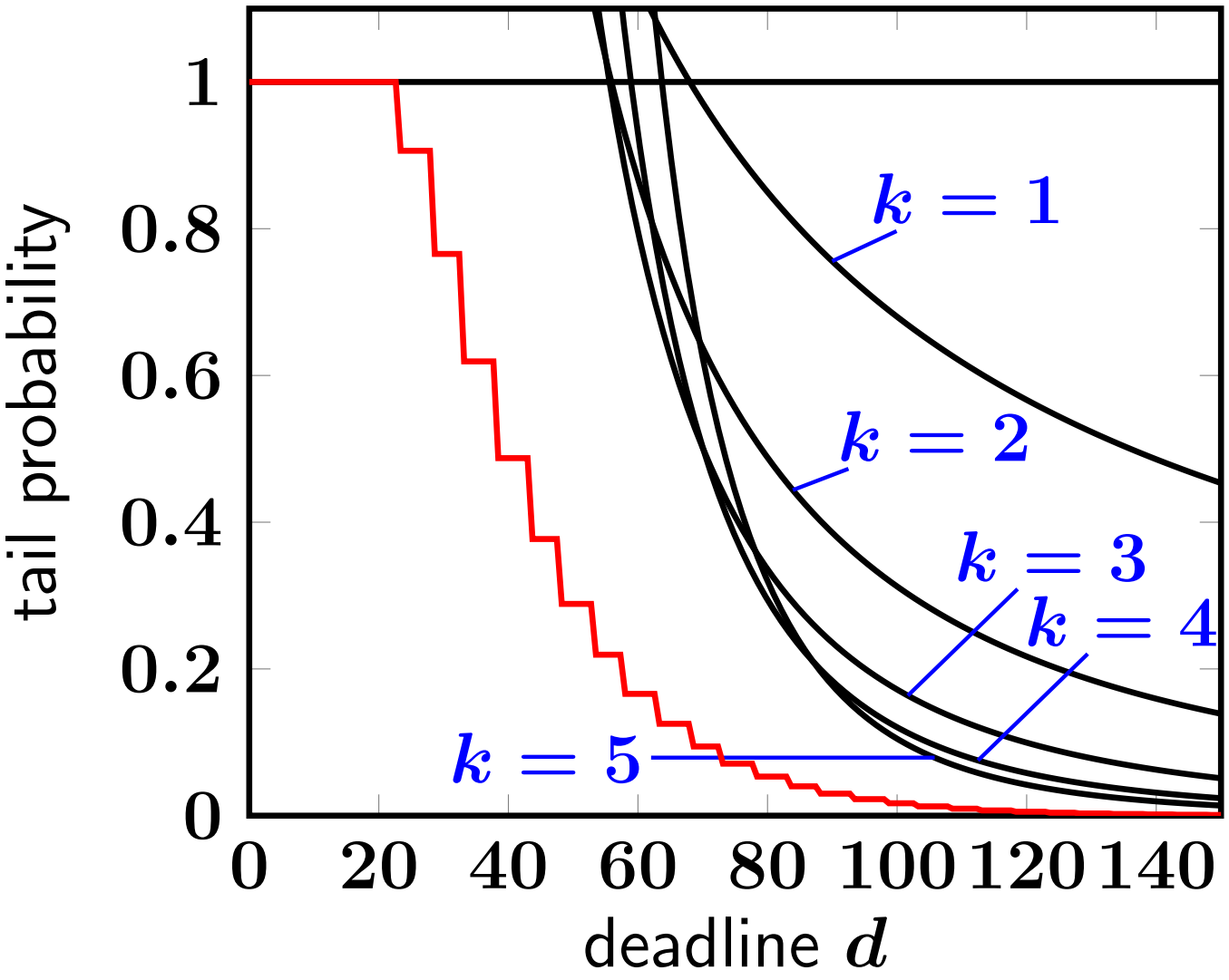
Experimental result (1)

A coupon collector's problem (linear template)

upper bound	execution time
$\mathbf{E}[T] \leq \mathbf{68}$	0.024 s
$\mathbf{E}[T^2] \leq \mathbf{3124}$	0.054 s
$\mathbf{E}[T^3] \leq \mathbf{171932}$	0.089 s
$\mathbf{E}[T^4] \leq \mathbf{12049876}$	0.126 s
$\mathbf{E}[T^5] \leq \mathbf{1048131068}$	0.191 s

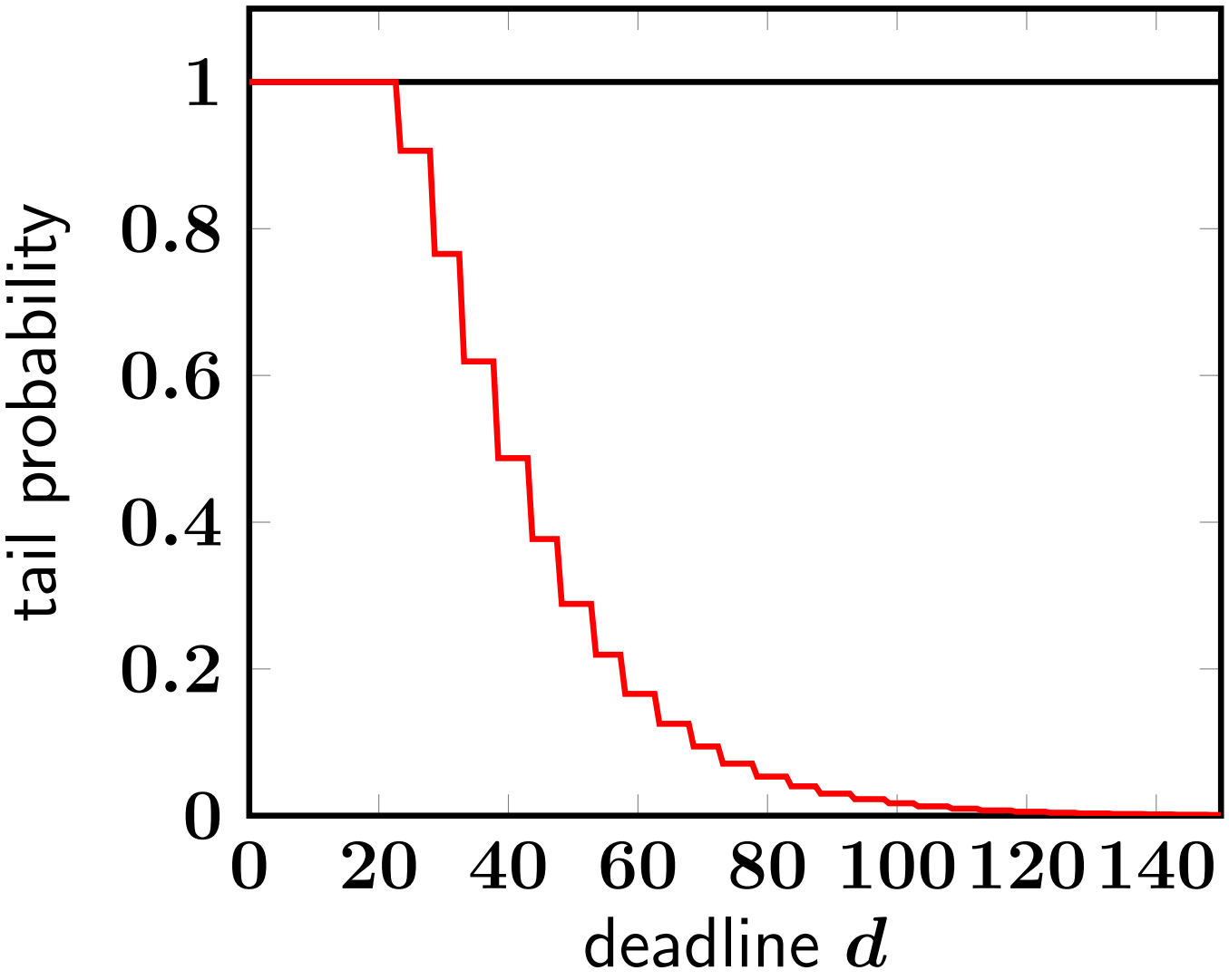
Experimental result (1)

$$\Pr(T \geq d) \leq \min_{k=0,\ldots,K} \frac{u_k}{d^k}$$



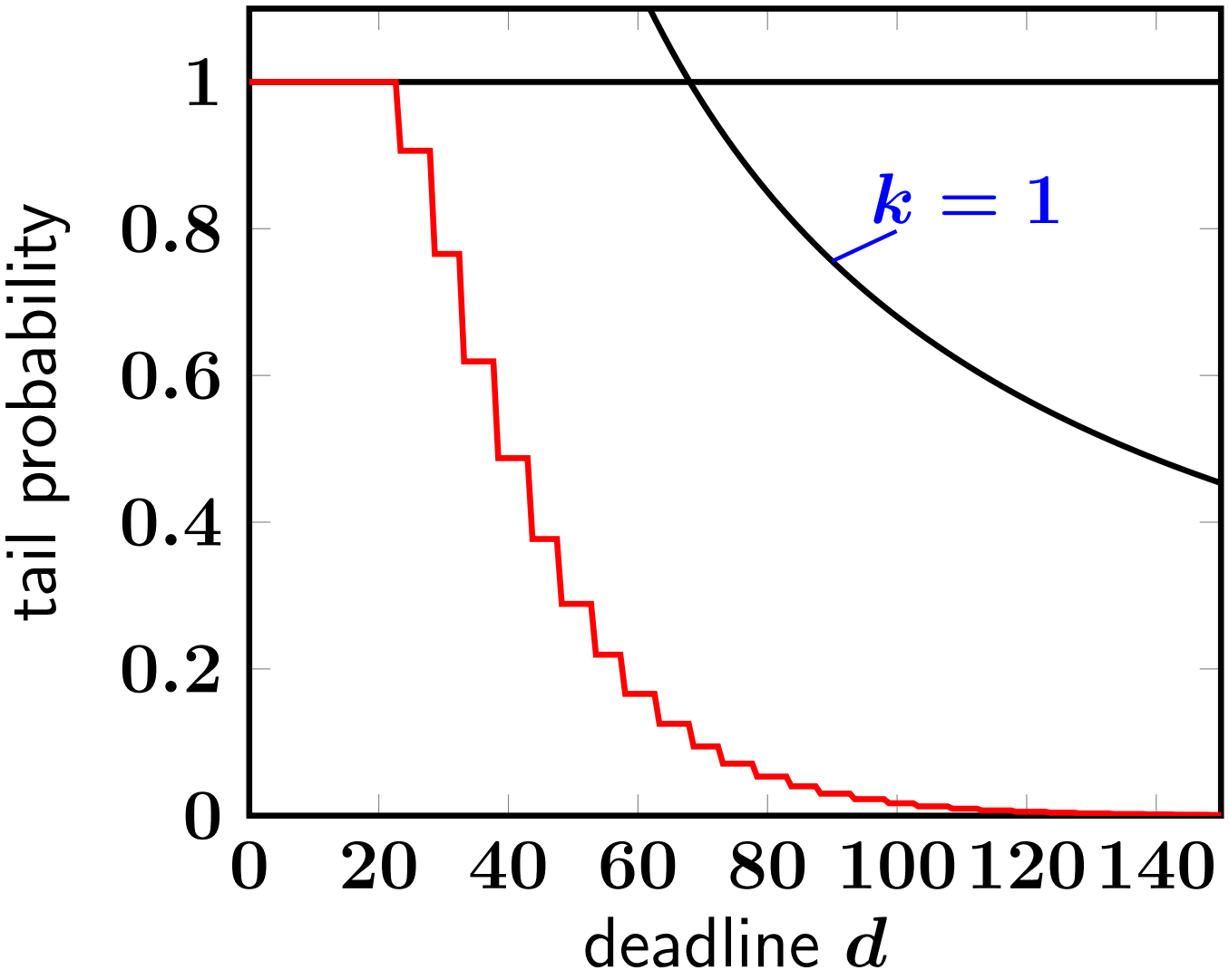
Experimental result (1)

$$\Pr(T \geq d) \leq \min_{k=0,\ldots,K} \frac{u_k}{d^k}$$



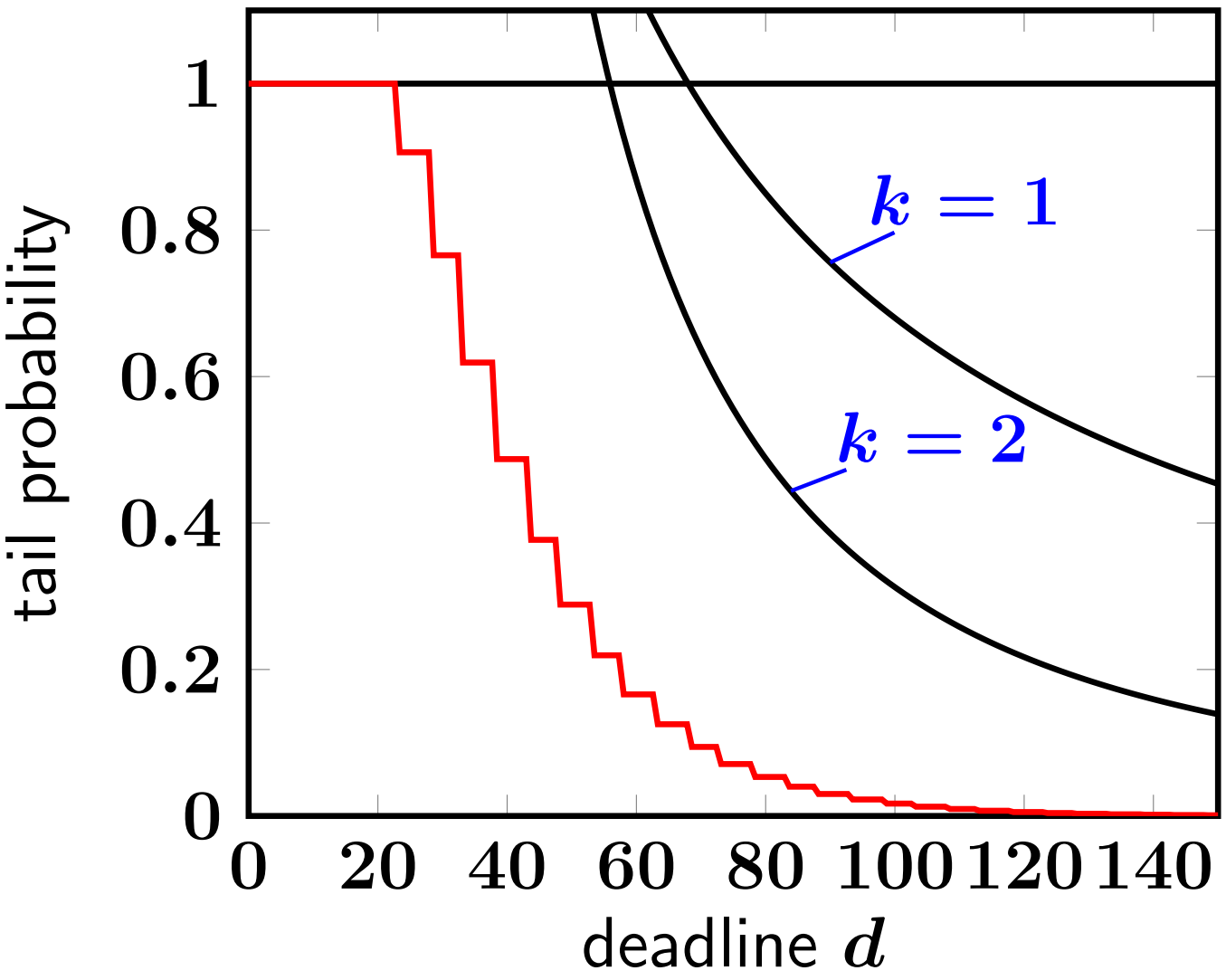
Experimental result (1)

$$\Pr(T \geq d) \leq \min_{k=0,\ldots,K} \frac{u_k}{d^k}$$



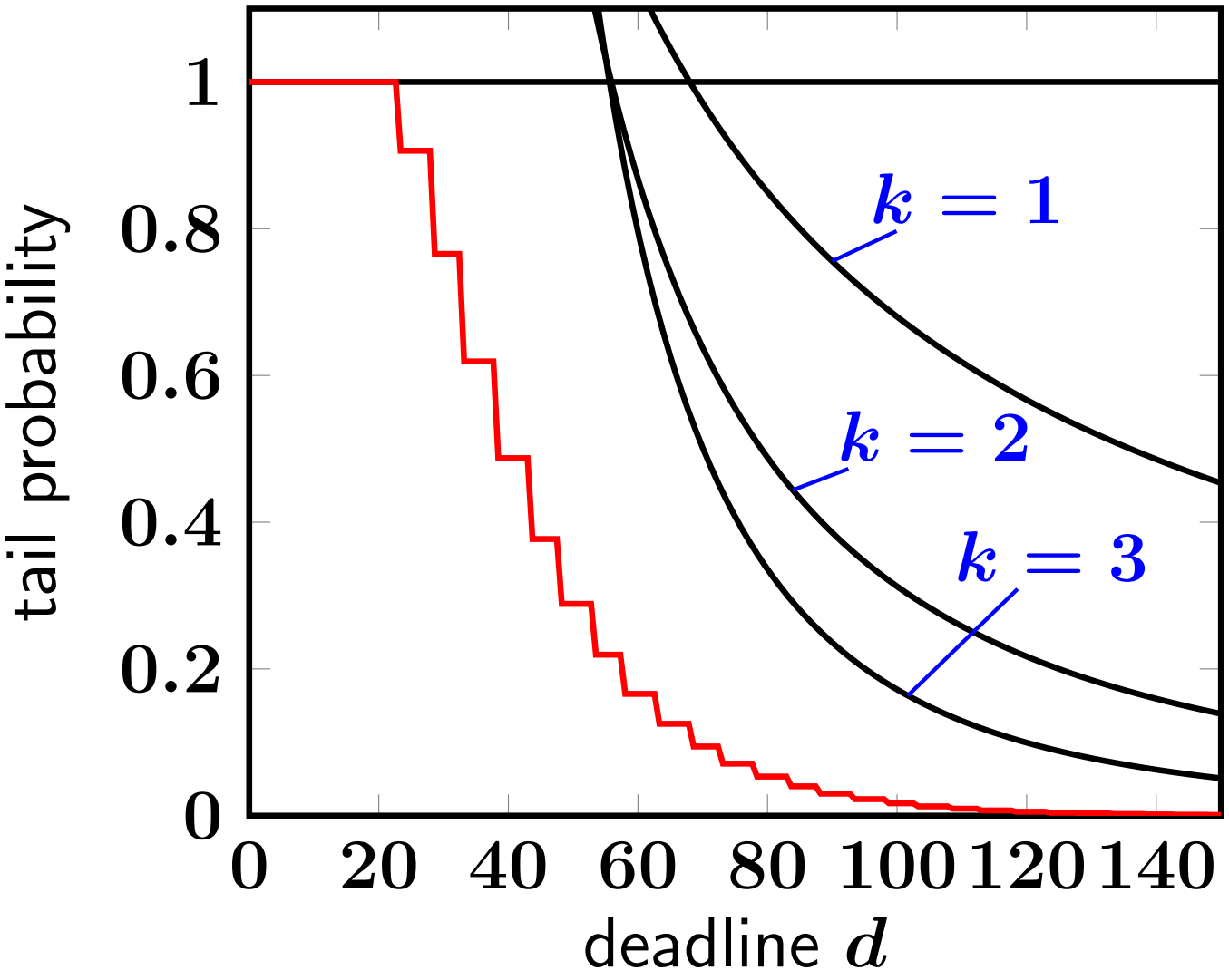
Experimental result (1)

$$\Pr(T \geq d) \leq \min_{k=0,\ldots,K} \frac{u_k}{d^k}$$



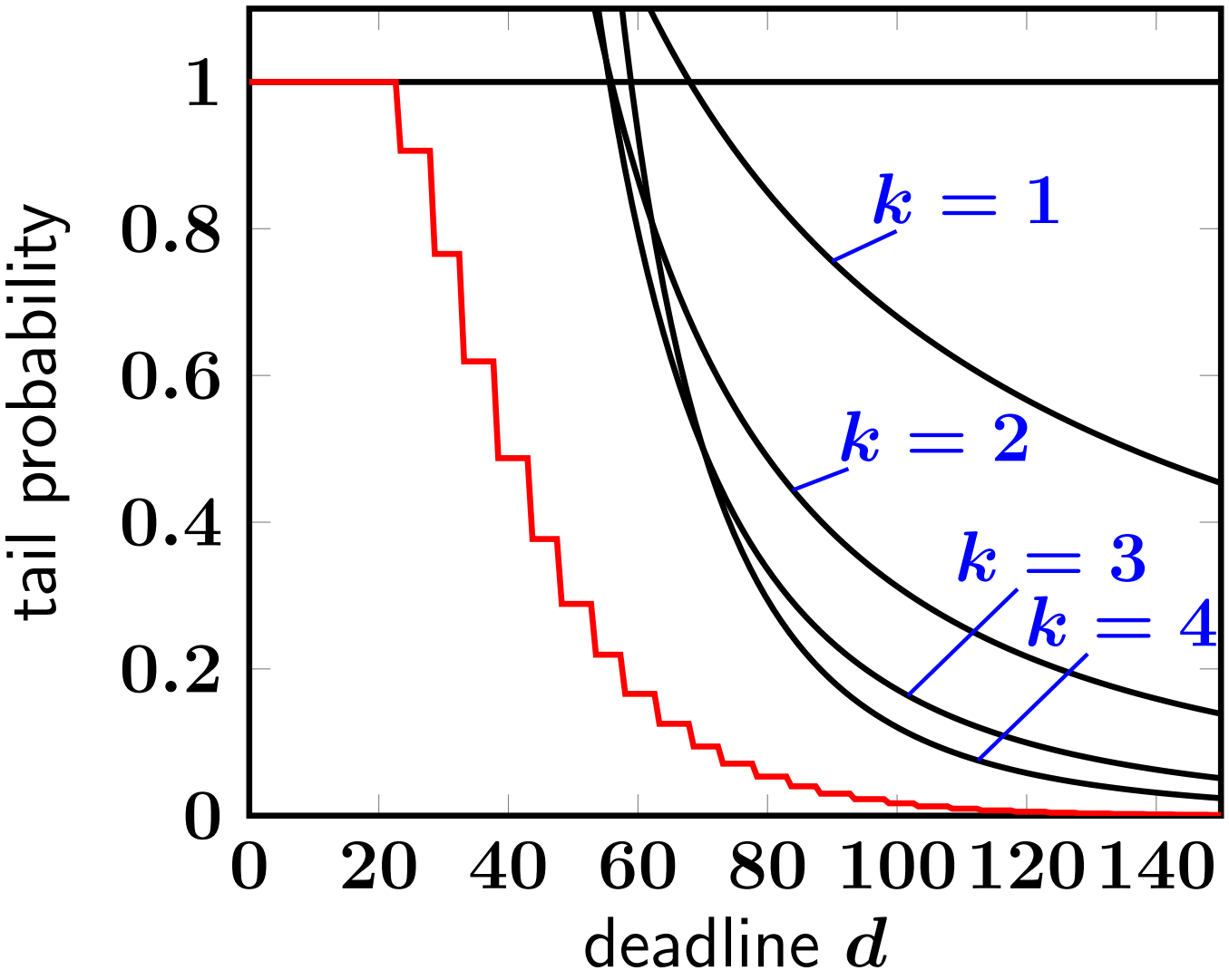
Experimental result (1)

$$\Pr(T \geq d) \leq \min_{k=0,\ldots,K} \frac{u_k}{d^k}$$



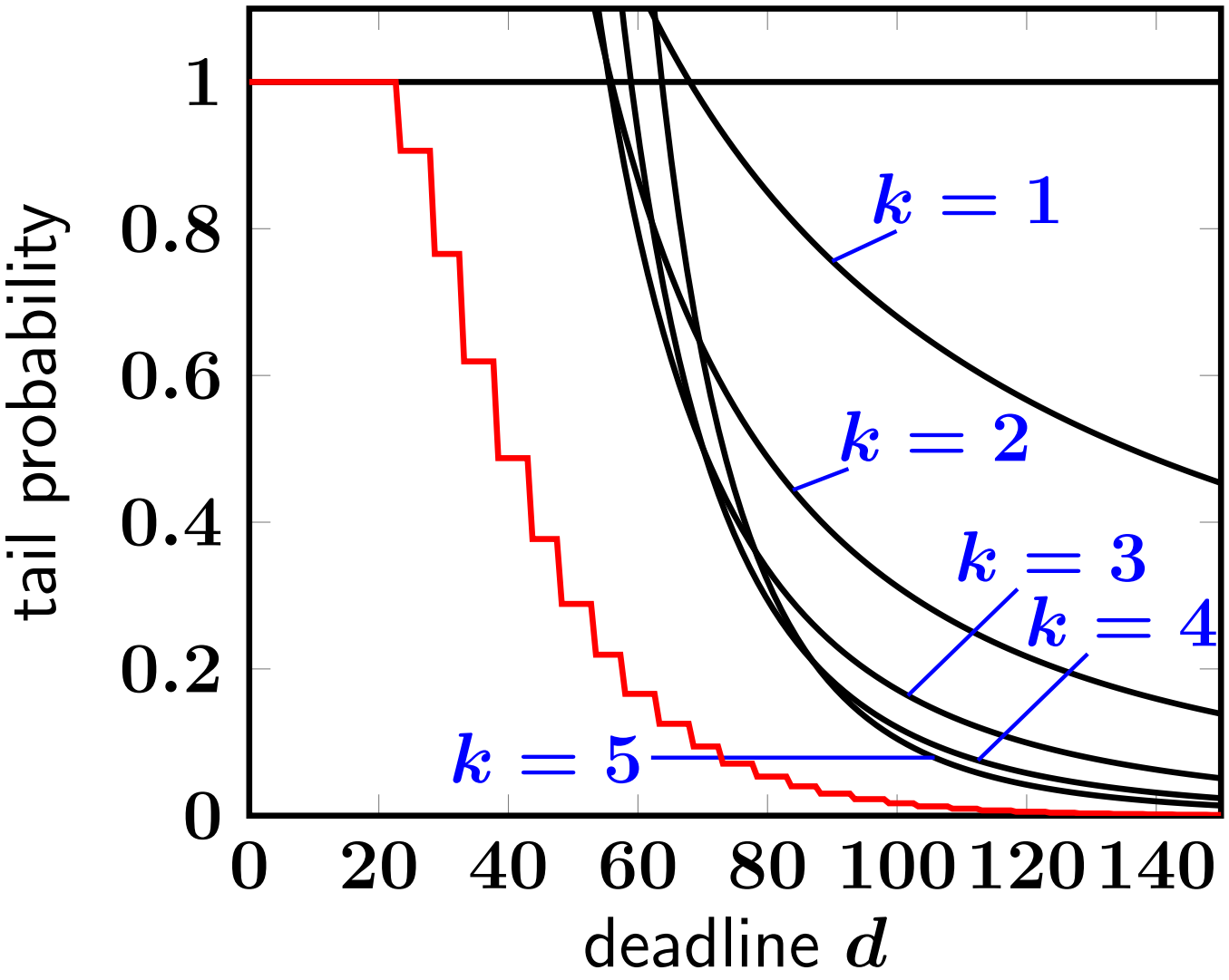
Experimental result (1)

$$\Pr(T \geq d) \leq \min_{k=0,\ldots,K} \frac{u_k}{d^k}$$



Experimental result (1)

$$\Pr(T \geq d) \leq \min_{k=0,\ldots,K} \frac{u_k}{d^k}$$



Experimental result (2)

A random walk with nondeterminism

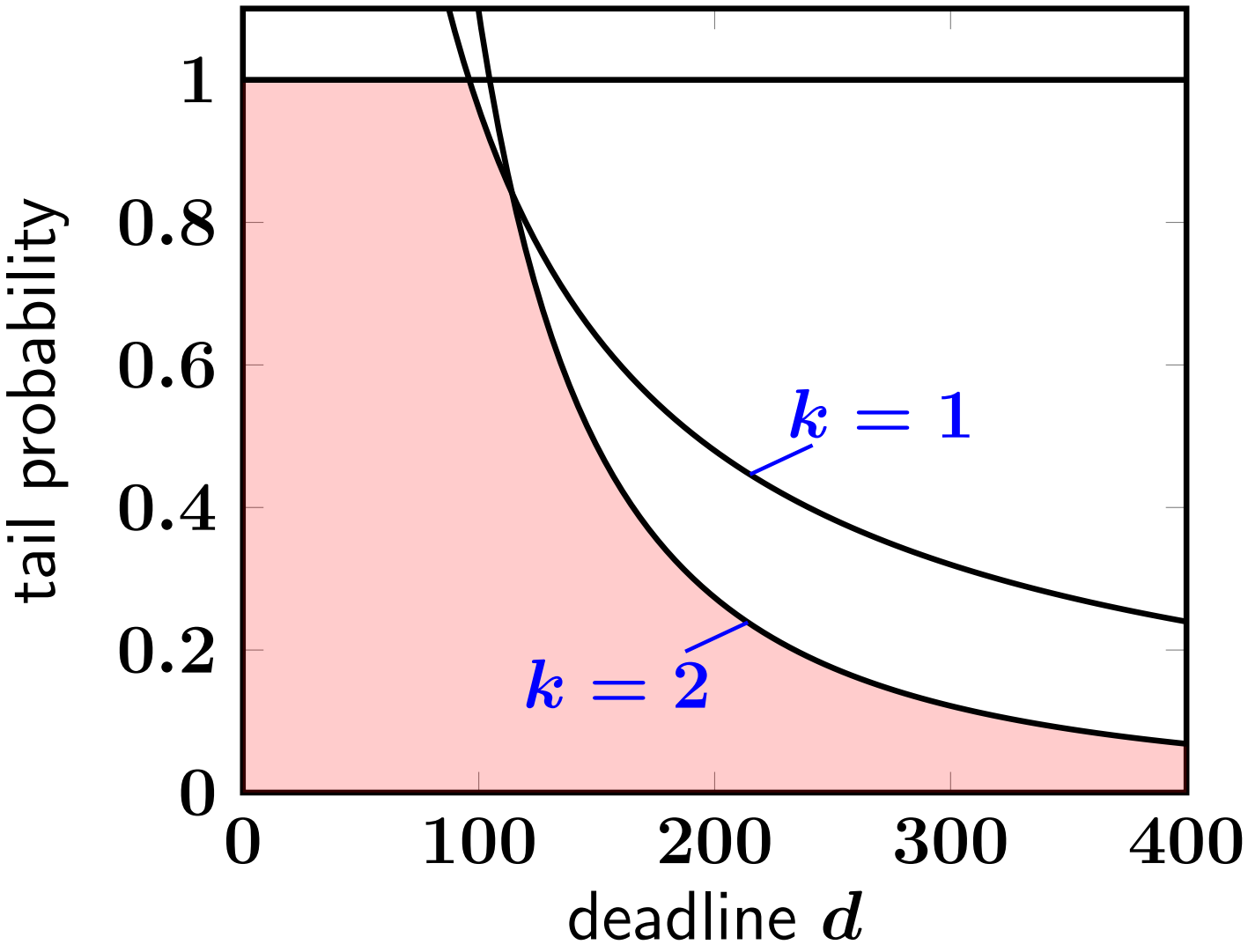
- Linear template

upper bound	execution time
$\mathbf{E}[\boldsymbol{T}] \leq \mathbf{96}$	0.020 s
$\mathbf{E}[\boldsymbol{T}^{\mathbf{2}}]$: infeasible	0.029 s

- Polynomial template

upper bound	execution time
$\mathbf{E}[\boldsymbol{T}] \leq \mathbf{95.95}$	157.748 s
$\mathbf{E}[\boldsymbol{T}^{\mathbf{2}}] \leq \mathbf{10944.0}$	361.957 s

Experimental result (2)



Conclusion & Future work

Conclusion

- New supermartingale for higher moments of runtime
- Applied to obtain upper bounds of tail probabilities
- Tested our method experimentally

Future work

- Improved treatment of nondeterminism
- Compositional reasoning (cf. [Kaminski et al., ESOP'16])
- Improve implementation (numerical error of SDP solver)