Importance Splitting for Statistical Model Checking Properties

Cyrille Jegourel, Axel Legay, Sean Sedwards

CAV 2013, Saint Petersburg

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Probabilistic model checking

Quantify temporal logical properties of stochastic systems

- *Numerical* model checking
  - precise
  - exhaustive exploration of state space
  - limited model size
- *Statistical* model checking (SMC)
  - statistical model of *executions*
  - results within confidence bounds
  - trades off tractability with precision

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Motivation

Objective :

- Standard Statistical technique for SMC: Monte Carlo.
- Rare events may cause serious problems and are difficult to simulate.
- Given a stochastic system, design a procedure for estimating a rare property in a reasonable time with SMC.
- Properties specified with time bounded temporal logic:

$$\phi = \alpha \mid \phi \vee \phi \mid \phi \wedge \phi \mid \neg \phi \mid \boldsymbol{X}\phi \mid \boldsymbol{F}^{t}\phi \mid \boldsymbol{G}^{t}\phi \mid \phi \boldsymbol{U}^{t} \phi$$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Monte Carlo model checking

Goal: Given a Markovian system and a property φ, compute the probability γ that a path ω satisfies φ ($\gamma = P[\omega \models \varphi]$).

The behavior of the system with respect to the property can be modeled by a Bernoulli random variable Z.

property indicator function $z \in \{0,1\}$

probability measure function

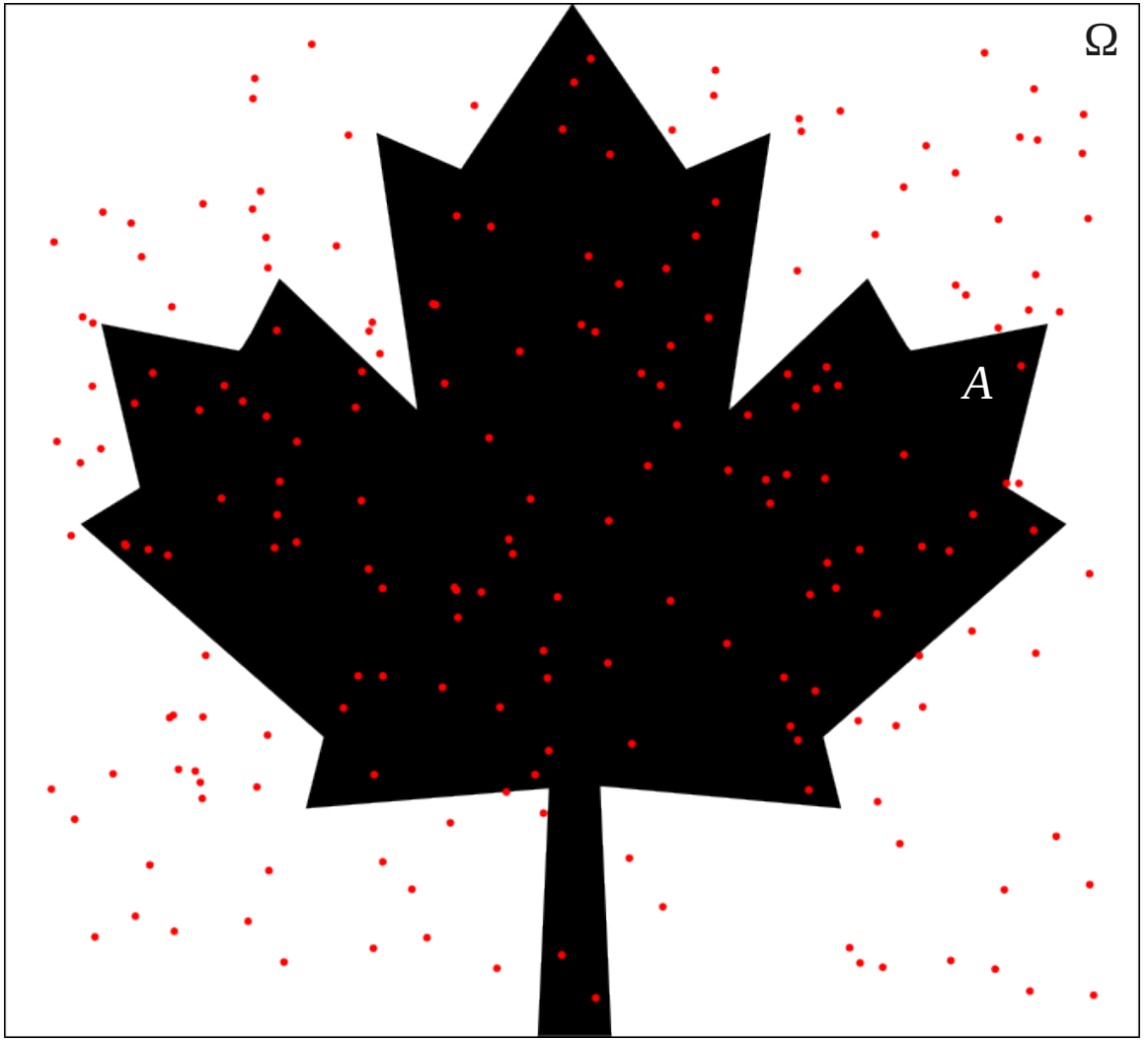
$$\gamma \stackrel{\text{def}}{=} \mathrm{E}_f[Z] = \int_{\Omega} z(\omega) f(\omega) d\omega$$

$$\tilde{\gamma} = \frac{1}{N} \sum_{i=1}^{N} z(\omega_i)$$

sample traces generated under *f*

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Monte Carlo estimation



$$A = \{\omega \in \Omega : z(\omega) = 1\}$$

$$\tilde{y} = \frac{1}{N} \sum_{i=1}^{N} z(\omega_i)$$

Absolute error = half the size of the confidence interval

$$\mathrm{AE} \propto \frac{\sqrt{\gamma(1-\gamma)}}{\sqrt{N}}$$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

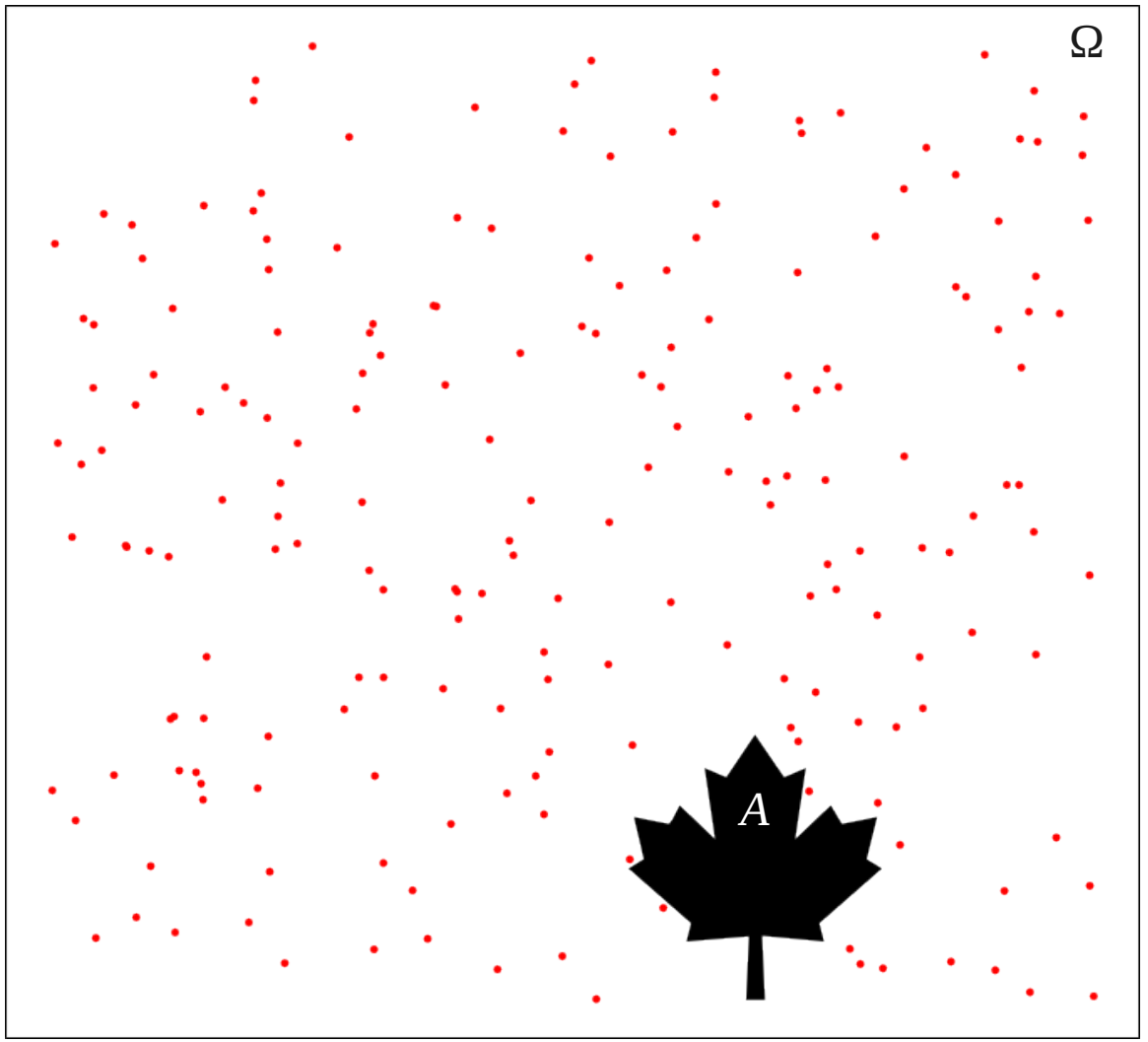
Problems of rare events

- Occur with small probability (e.g. $< 10^{-6}$)
  - appear rarely in stochastic simulations
  - need very large number of trials to see single example
  - without seeing, cannot quantify how low the probability
- The absolute error is not useful $(\gamma \pm \epsilon)$ not useful if $\epsilon \gg \gamma$
  - Bounds (e.g. Chernoff) not useful when γ small
  - Unbounded *relative* error:

$$\mathrm{RE} = \frac{\sqrt{Var(z)}}{E(z)} = \frac{\sqrt{\gamma - \gamma^2}}{\gamma} \approx_{\gamma \to 0} \frac{1}{\sqrt{\gamma}}$$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

High variance



$$\mathrm{RE} \propto_{\gamma \to 0} \frac{1}{\sqrt{N\,\gamma}}$$

N very large to bound RE
with Monte Carlo simulation

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Importance sampling

Monte Carlo

$$\gamma = \int_{\Omega} z(\omega)\, f(\omega)\, d\omega$$

$$\tilde{\gamma}_{MC} = \frac{1}{N} \sum_{i=1}^{N} z(\omega_i)$$

traces generated under f

$$\gamma = \int_{\Omega} z(\omega) \frac{f(\omega)}{f'(\omega)} f'(\omega)\, d\omega$$

importance sampling distribution

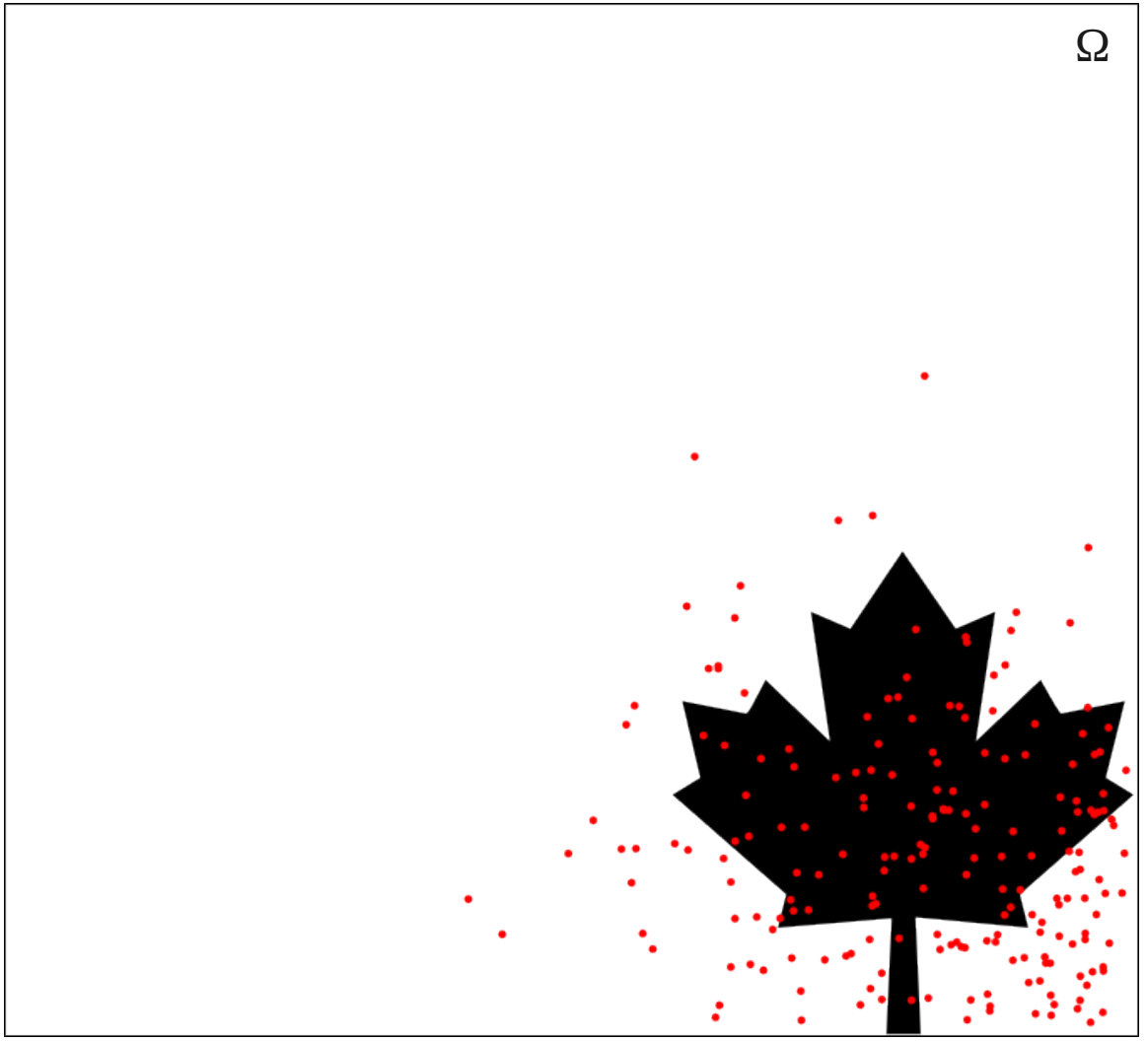
likelihood ratio

$$\tilde{\gamma}_{IS} = \frac{1}{N} \sum_{i=1}^{N} z(\omega_i) \frac{f(\omega_i)}{f'(\omega_i)}$$

traces generated under f'

Inria
INVENTEURS DU MONDE NUMÉRIQUE

‘Tilted’ simulation

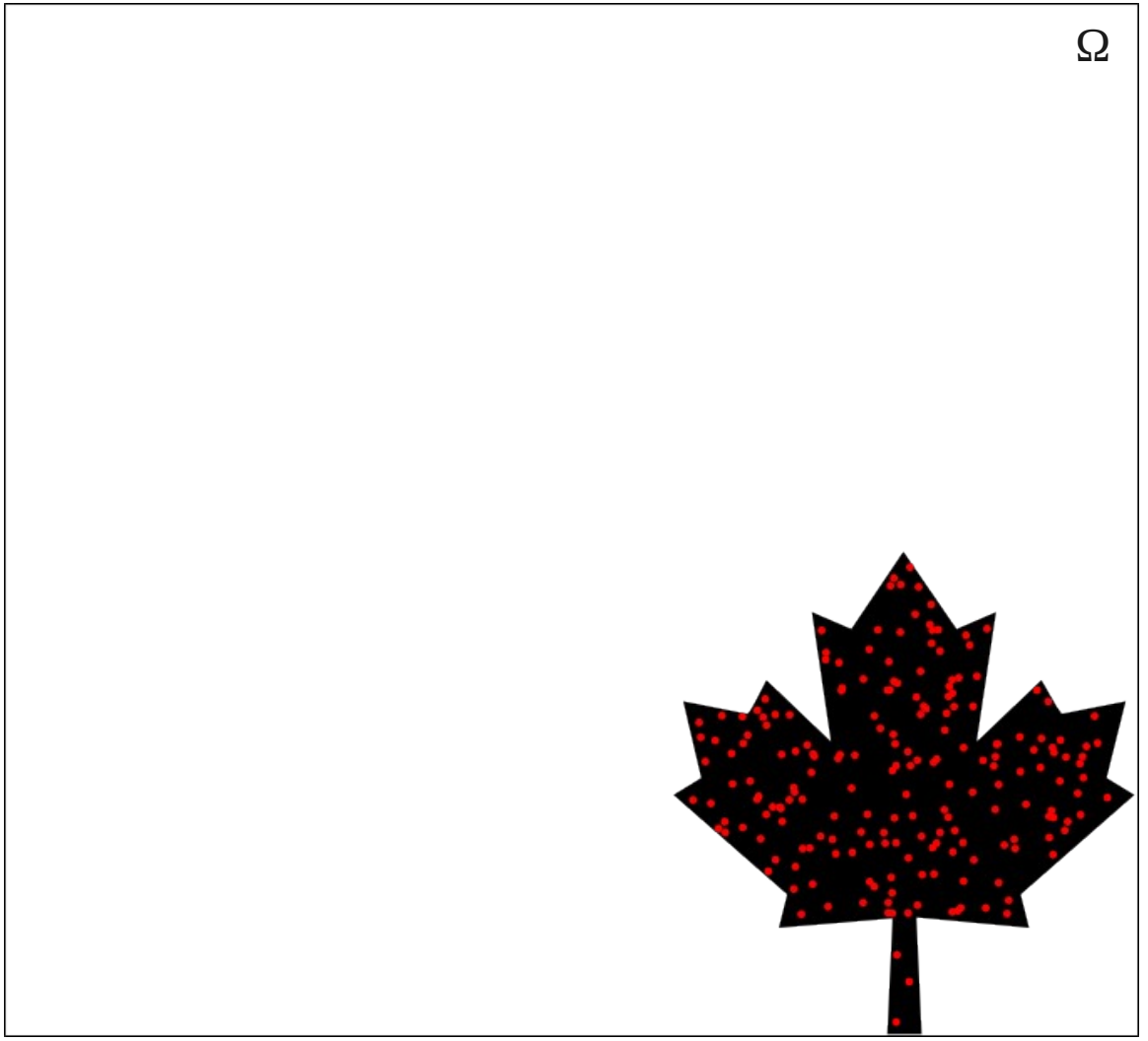


$$\tilde{y} = \frac{1}{N} \sum_{i=1}^{N} z(\omega'_i) \frac{f(\omega')}{f'(\omega'_i)}$$

traces generated under f'
(importance sampling dist.)

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Optimal importance sampling



$$\tilde{y} = \frac{1}{N} \sum_{i=1}^{N} z(\omega'_i) \frac{f(\omega')}{f'(\omega'_i)}$$

$$f^{opt} = \frac{z f}{y}$$

f conditioned on the rare event

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Limitations of Importance Sampling

- Quantifying the performance of apparently “good” distributions is an open problem.
- Problem of accuracy with long simulations: likelihood ratio vanishes and variance of the estimators increases.

=> need of an alternative technique: Importance Splitting.

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Basics of Importance Splitting

Let A a rare event and $(A_k)_{0 \le k \le n}$ a sequence of nested events:

$$A_0 \supset A_1 \supset \ldots \supset A_n = A$$

$$\gamma \stackrel{\text{def}}{=} P(A) = P(A_0) P(A_1 | A_0) P(A_2 | A_1) \ldots P(A_n | A_{n-1})$$

Bayes formula

$$\forall k \quad P(A_k \mid A_{k-1}) = \gamma_k \ge \gamma$$

Less rare

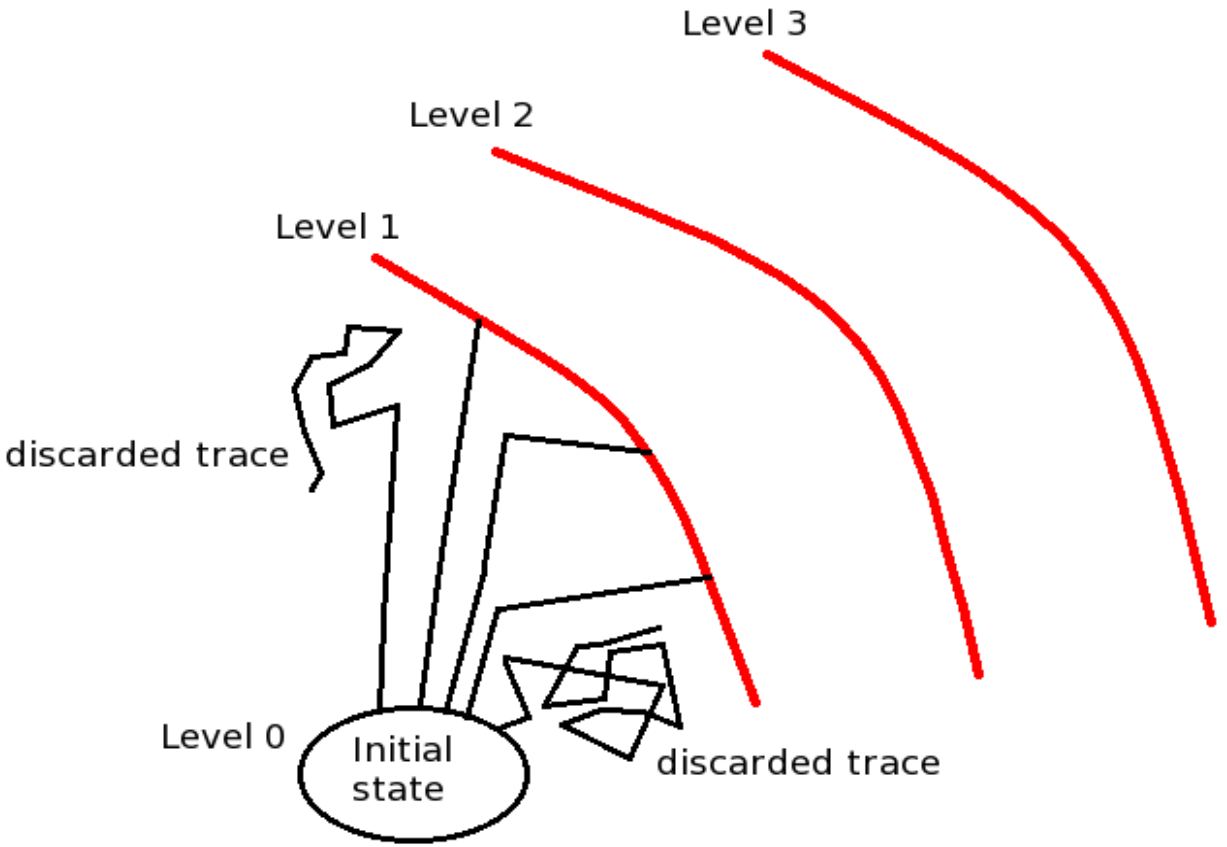
Inria
INVENTEURS DU MONDE NUMÉRIQUE

Generation of traces in Importance Splitting

- Assuming a set of increasing *levels k*, generate traces starting from a distribution of the initial states.
- Simulations are stopped as soon as they reach the next level k+1.
- The final states become the empirical distribution of initial states for the next level (level k+1).
- Failed traces discarded. Successful traces continue from where they stopped.
- Avoid a reduction of simulations by resampling the discarded traces from empirical distribution of level k+1.

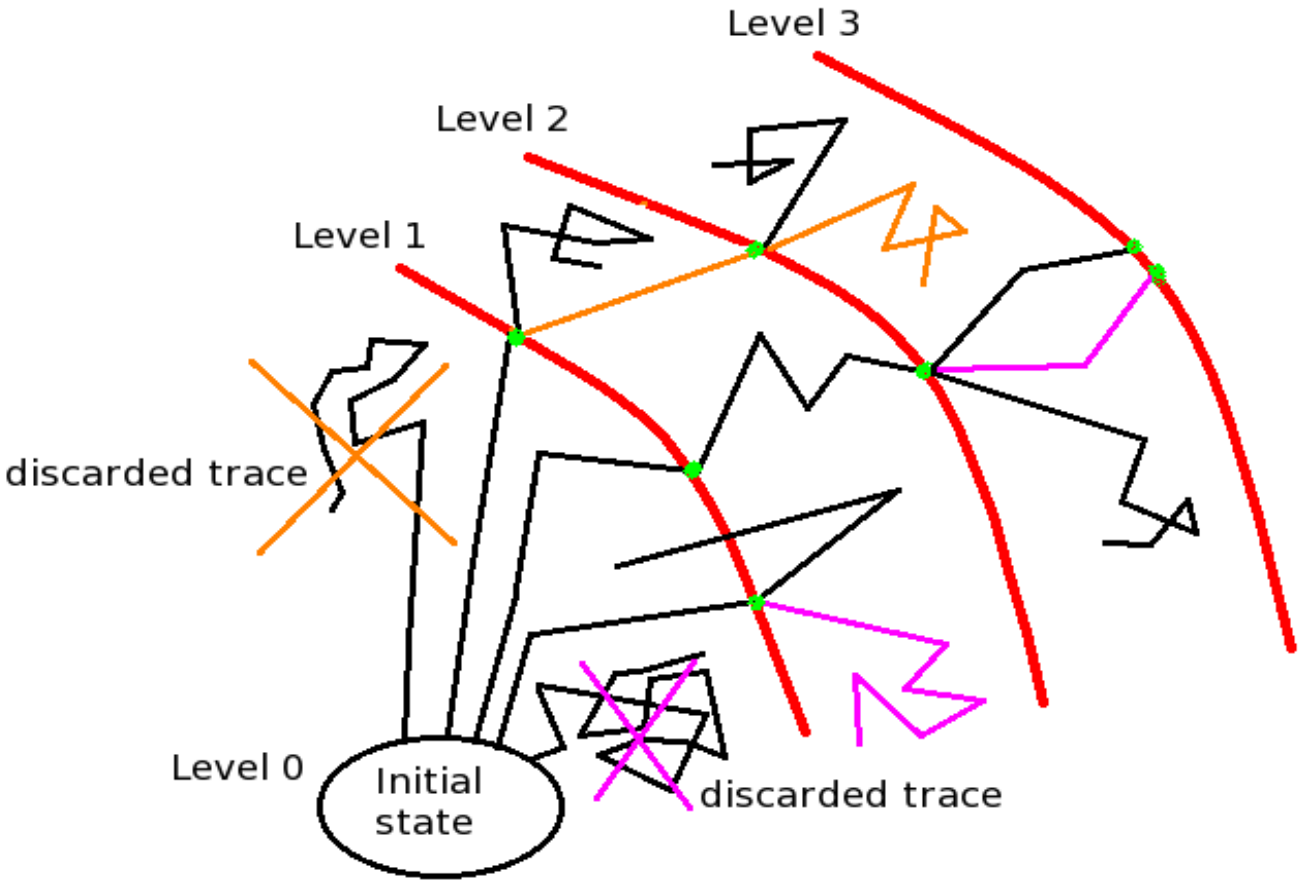
Inria
INVENTEURS DU MONDE NUMÉRIQUE

Illustration 1



Inria
INVENTEURS DU MONDE NUMÉRIQUE

Illustration 2



P(reaching Level 3)=3/5*2/5*2/5

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Importance Splitting for (Temporal) Logic

Idea: given a rare property ϕ, define a set of levels based on a sequence of temporal properties such that:

$$(\phi_k)_{0 \le k \le n} : \ \phi_0 \Leftarrow \phi_1 \Leftarrow \ldots \Leftarrow \phi_n = \phi$$

Thus,

$$\gamma = P(\omega \vDash \phi_0) \prod_{k=1}^{n} P(\omega \vDash \phi_k \mid \omega \vDash \phi_{k-1})$$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Level-based Score functions

- Goal: Generalise the concept of levels.
- Definition 1:

  Let $J_0 \supset J_1 \supset ... \supset J_n$ be a set of nested intervals of $\mathbb{R}$.

  Let $\phi_0 \supset \phi_1 \supset ... \supset \phi_n$ be a set of nested properties.

  $S : \Omega \to \mathbb{R}$ is a level-based score function of property ϕ iff $\forall k$:

$$\omega \models \phi_k \Leftrightarrow S(\omega) \in J_k \text{ and } \forall i, j \in \{0, ..., |\omega|\}: i < j \Rightarrow S(\omega_{\leq i}) \leq S(\omega_{\leq j})$$

- Example: given a set of nested properties, a simple score function may be defined as follows:

$$S(\omega) = \sum_{k=1}^{n} 1(\omega \models \phi_k)$$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

General score functions

- Definition 2:

  Let $J_0 \supset J_1 \supset \ldots \supset J_n$ be a set of nested intervals of $\mathbb{R}$.

  Let $\Omega = \Omega_0 \supset \Omega_1 \supset \ldots \supset \Omega_n$ be a set of nested subsets of Ω.

  $S : \Omega \to \mathbb{R}$ is a general score function of property ϕ iff $\forall k$:

  (i) $\omega \in \Omega_k \Leftrightarrow S(\omega) \in J_k$

  (ii) $\omega \vDash \phi \Leftrightarrow \omega \in \Omega_n$

  (iii) $\forall i, j \in \{0, \ldots, |\omega|\}: i < j \Rightarrow S(\omega_{\leq i}) \leq S(\omega_{\leq j})$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Score functions

- Goal: Generalise the concept of levels.
- Level-based score functions: Mapping from logical properties to the real numbers which give information on the number of satisfied sub-formulae.

$$\text{Example: } S(\omega) = \sum_{k=1}^{n} 1(\omega \models \phi_k)$$

- General score functions: Mapping from sets of paths to the real numbers s.t. higher scores assigned to paths that satisfy the overall property.

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Use of Heuristics

- Level-based score functions correlate logic to score.
- General score functions requires:
  - higher scores assigned to paths that satisfy the overall property.
  - Score of a path's prefix is non decreasing with increasing prefix length.
- In some case, the shortest paths satisfying a rare property are the most likely.

  => possibility to exploit the length of a path to improve a score function based on coarse logical levels.

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Simple decomposition

- When $\phi = \Lambda_{j=1}^{n} \psi_j$, a decomposition into nested properties is:

$$\phi_i = \Lambda_{j=1}^{i} \psi_j, \quad \forall i \in \{1, \ldots, n\} \text{ with } \phi_0 \equiv True$$

- Possibility to choose an arbitrary order of sub-formulae:

Example: Given $\phi = a \wedge b \wedge c$,

$$\phi_3 = a \wedge b \wedge c, \ \phi_2 = a \wedge b, \ \phi_1 = a$$

$$\phi_3 = a \wedge b \wedge c, \ \phi_2 = b \wedge c, \ \phi_1 = c$$

Both decompositions are valid.

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Natural decomposition

- Many rare events are defined with a natural notion of level, when some quantity of the system reaches a particular value.
- In Computational systems: might refer to a loop counter, a number of software objects, etc...
- In physical systems: might refer to a temperature, a distance, a number of molecules...
- Natural levels defined by nested atomic properties:

$$\phi_i = (l > l_i) \text{ with } l \text{ a state variable and } \omega \models \phi_n \Leftrightarrow l \geq l_n$$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Decomposition of temporal operators

$(i)\ \ (\phi_n \Rightarrow \phi_{n-1}) \Rightarrow (\boldsymbol{S}\phi_n \Rightarrow \boldsymbol{S}\phi_{n-1})$ with $\boldsymbol{S} \in \{\boldsymbol{F}^{\leq t}, \boldsymbol{G}^{\leq t}, \boldsymbol{X}, \boldsymbol{F}^{\leq t}\boldsymbol{G}^{\leq s}\}$

$(ii)\ \ (\phi_n \Rightarrow \phi_{n-1} \wedge \psi_m \Rightarrow \psi_{m-1}) \Rightarrow (\phi_n \boldsymbol{U} \psi_m \Rightarrow \phi_{n-1} \boldsymbol{U} \psi_{n-1})$

$(iii)\ \ (\phi_n \Rightarrow \phi_{n-1}) \Rightarrow (\forall \omega \vDash \boldsymbol{G}^{\leq t}\phi_n : \exists t' \geq t \mid \omega \vDash \boldsymbol{G}^{\leq t'}\phi_{n-1})$

$(iv)\ \ (\phi_n \Rightarrow \phi_{n-1}) \Rightarrow (\forall \omega \vDash \boldsymbol{F}^{\leq t}\phi_n : \exists t' \leq t \mid \omega \vDash \boldsymbol{F}^{\leq t'}\phi_{n-1})$

$(v)\ \ (t' \geq t \wedge s' \leq s) \Rightarrow (\boldsymbol{F}^{\leq t}\boldsymbol{G}^{\leq s}\phi_n \Rightarrow \boldsymbol{F}^{\leq t'}\boldsymbol{G}^{\leq s'}\phi_n)$

$(vi)\ \ (\phi_n \Rightarrow \phi_{n-1}) \Rightarrow (\forall \omega \vDash \boldsymbol{F}^{\leq t}G^{\leq s}\phi_n : \exists t' \leq t \wedge s' \geq s \mid \omega \vDash \boldsymbol{F}^{\leq t'}\boldsymbol{G}^{\leq s'}\phi_{n-1})$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Two algorithms

- Fixed level algorithm:
  - Exploits a score function based on “logical” levels

- Adaptive level algorithm:
  - Given a score function, finds itself the “best” levels
  - Requires a score function refined enough.

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Fixed level algorithm

Let $(\tau_k)_{1 \le k \le M}$ be the sequence of thresholds

Let **stop** be a termination condition

for $1 \le k \le M$ **do**

$\forall\, 1 \le j \le N$, using prefix $\tilde{\omega}_j^k$, generate path ω_j^k until $S(\omega_j^k) \ge \tau_k \vee$ **stop**

$I_k = \{\ j : S(\omega_j^k) \ge \tau_k\ \}$ and $\tilde{\gamma}_k = \dfrac{|I_k|}{N}$

$\forall\, j \in I_k,\ \tilde{\omega}_j^{k+1} = \omega_j^k$

$\forall\, j \notin I_k$, let $\tilde{\omega}_j^{k+1}$ be a copy of ω_i^k with $i \in I_k$ chosen randomly

$$\tilde{\gamma} = \prod_{k=1}^{M} \tilde{\gamma}_k$$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Fluctuation analysis

- Estimator unbiased
- Confidence interval based on the relative variance:

$$\sqrt{N}\frac{\tilde{\gamma}-\gamma}{\gamma} \underset{n\to\infty}{\overset{D}{\to}} \boldsymbol{N}(0,\sigma^2) \text{ where } \boldsymbol{N} \text{ denotes a Gaussian distribution}$$

$$\text{with } \sigma^2 \geq \sum_{k=0}^{n-1} \frac{1-\gamma_k}{\gamma_k}$$

- Inequality arises because the independence of initial states diminishes with increasing levels.
- Several possibilities minimise this dependence effect.

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Idealized version

- relative variance of the estimator: $\sigma^2 = \sum_{k=0}^{n-1} \frac{1-\gamma_k}{\gamma_k}$

- For a fixed number of levels, this variance is minimal if all the conditional probabilities are equal:

$$\underset{p_0,\cdots p_{n-1}}{argmin} \sum_{k=0}^{n-1} \frac{1-\gamma_k}{\gamma_k} \quad \text{s.t.} \quad \prod_{k=0}^{n-1} \gamma_k = p$$

- It corresponds to the case where the levels are evenly spaced in terms of probability of success.
- Hence, the idea of an adaptive algorithm.

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Adaptive level algorithm

Let N_k be the predefined number of paths to keep per iteration

Let τ_ϕ be the minimum score of paths that satisfy ϕ

$k=1$. $\forall\, 1 \le j \le N$, generate path ω_j^k

repeat (until $\tau_k \ge \tau_\phi$)

Let $T = \{ S(\omega_j^k),\ \forall\, j \in \{1, \ldots, N\} \}$

Find minimum $\tilde{\tau}_k$ s.t. $|\{\tau \in T : \tau > \tilde{\tau}_k\}| \ge N_k$

$I_k = \{\ j : S(\omega_j^k) \ge \tau_k\ \}$ and $\tilde{\gamma}_k = \dfrac{|I_k|}{N}$

$\forall\, j \in I_k,\ \omega_j^{k+1} = \omega_j^k$

for $j \notin I_k$ **do**

Choose randomly $l \in I_k$

$\tilde{\omega}_j^{k+1} = max\{\omega \in pref(\omega_l^k) : S(\omega) < \tau_k\}$

generate path ω_j^{k+1} with prefix $\tilde{\omega}_j^{k+1}$

$M, k = k, k+1$

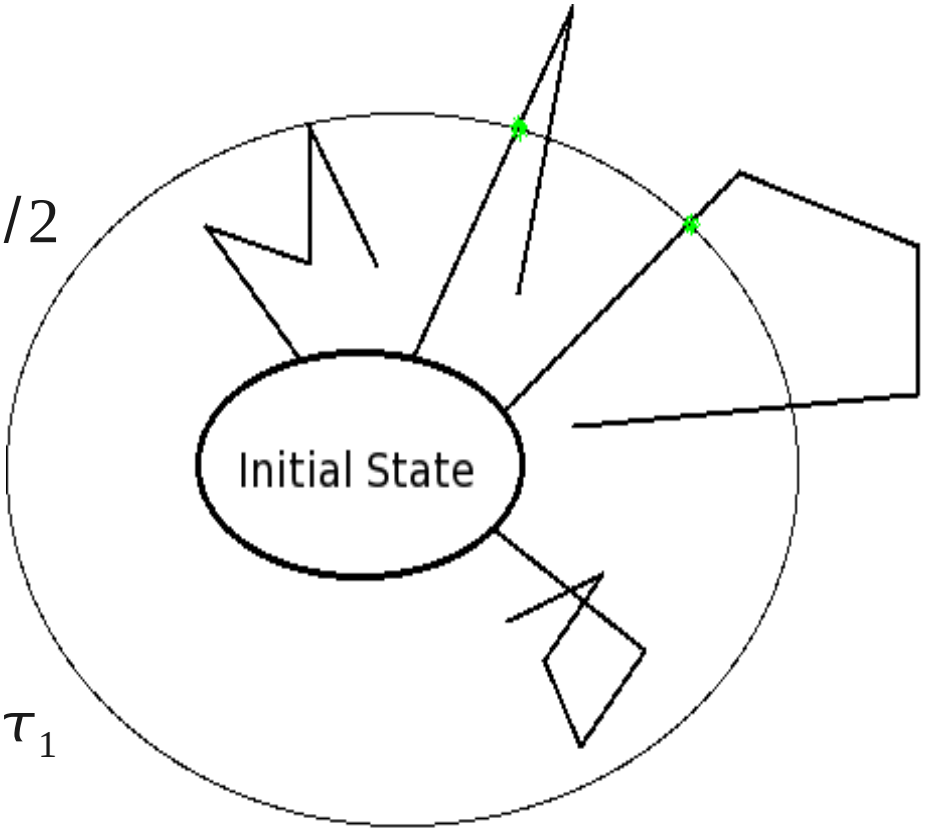
$$\tilde{\gamma} = \prod_{k=1}^{M} \tilde{\gamma}_k$$

$^{|\omega|}$

Adaptive algorithm illustration
1st step: trace generation

Predefine $\gamma_0 = 1/2$

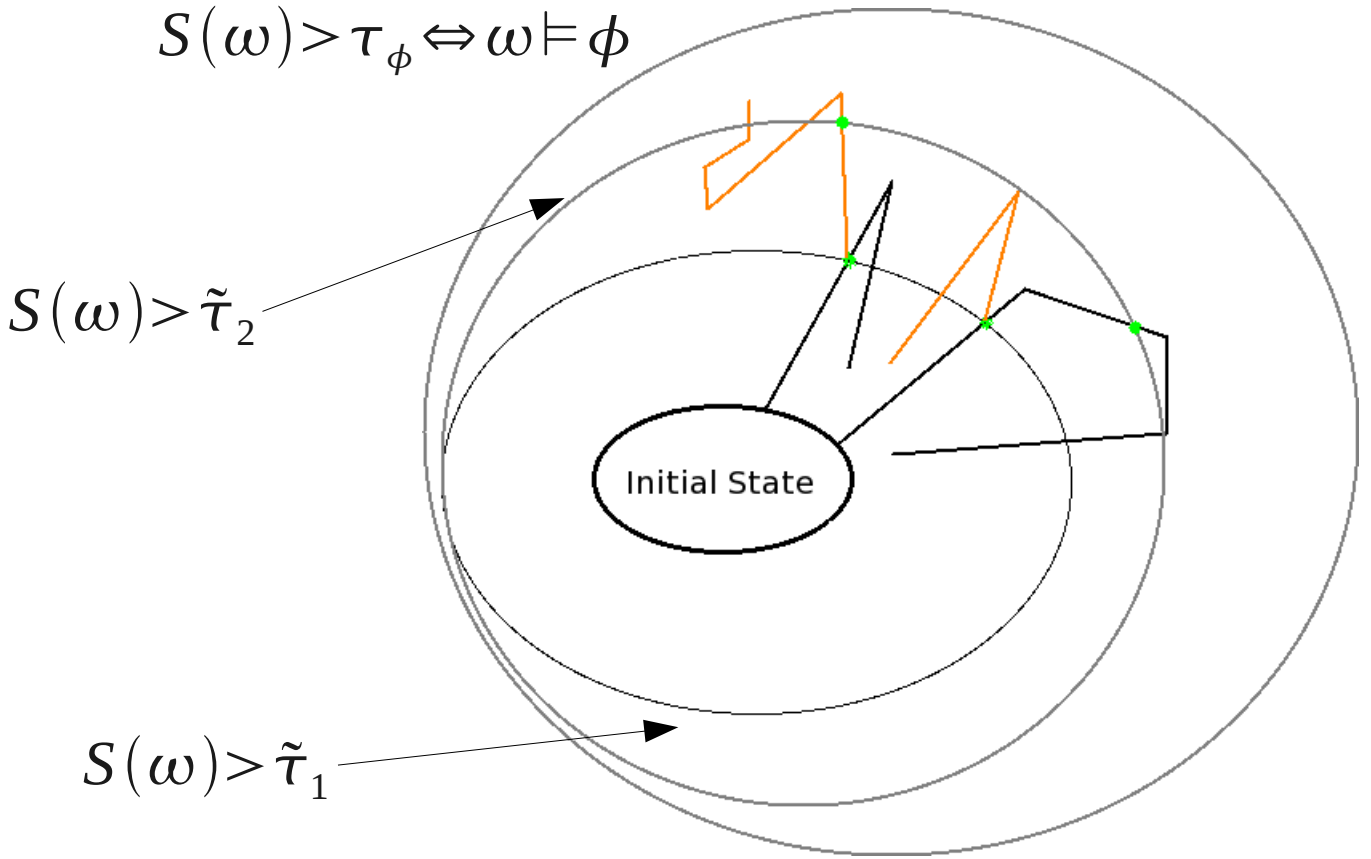
Estimate τ_1 s.t. $P(Score(\omega) > \tau_1) = 1/2$



Ellipse corresponding to $Score(\omega) = \tau_1$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Adaptive algorithm illustration



Inria
INVENTEURS DU MONDE NUMÉRIQUE

Fluctuation analysis

- For simplicity, let us write: $\gamma = r\,\gamma_0^{\,M}$

  with γ_0 the predefined conditionnal probability,

  M the number of levels

  r the number of traces ω in the last iteration s.t.: $S(\omega) \geq \tau_\phi$

$$\sqrt{N}\,\frac{\tilde{\gamma}-\gamma}{\gamma} \underset{n\to\infty}{\overset{D}{\to}} \mathbf{N}(0,\sigma^2) \qquad \text{with } \sigma^2 \geq M\,\frac{1-\gamma_0}{\gamma_0} + \frac{1-r}{r}$$

$$E[\tilde{\gamma}] - \gamma \sim \frac{\gamma}{N}\,\frac{M(1-\gamma_0)}{\gamma_0}$$

- Positive Bias of order O(1/N) => Good news!

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Bias, variance and Confidence interval

- $E[\tilde{y}] = p(1+O(N^{-1}))$
- $Var(\tilde{y}) = \frac{p^2}{N}(M\frac{1-\gamma_0}{\gamma_0}+\frac{1-r}{r})+o(N^{-1})$
- => Hence, the variance is reduced if γ_0 is chosen large.
- Confidence interval of level (1-α) based on the relative variance:

$$[\ \tilde{y}(\frac{1}{1+z_\alpha\sigma N^{-1/2}}),\ \tilde{y}(\frac{1}{1-z_\alpha\sigma N^{-1/2}})$$

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Example: Dining philosophers

Model: 150 philosophs

Property of interest:

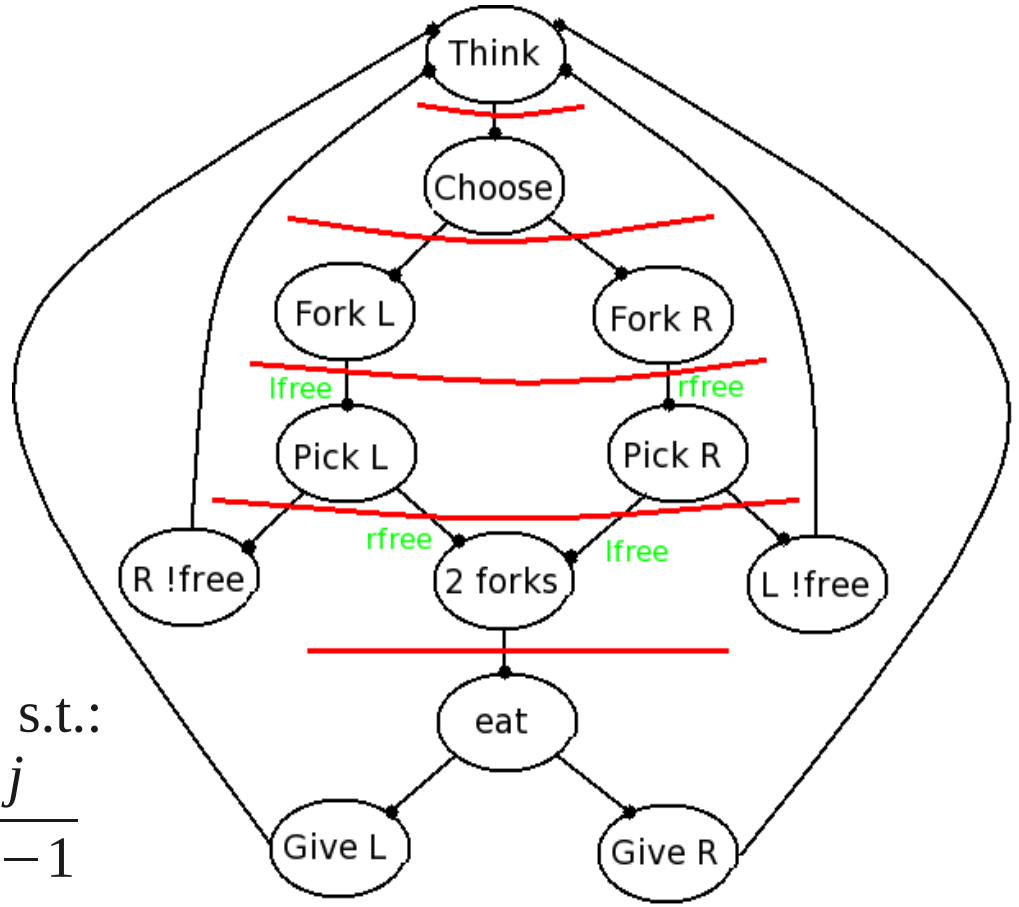
$$\boldsymbol{F}^{\leq K}\ \text{phil}_i\ \textit{eat}$$

$$S_1(\omega) = \sum_{k=1}^{n} 1\,(\omega \vDash \phi_k)$$

$$S_2(\omega) = \max_{1 \leq j \leq K} \Psi\,(\omega_{\leq j})$$

where j is prefix's length and Ψ s.t.:

$$\Psi\,(\omega_{\leq j}) = S_1(\omega_{\leq j}) - \frac{S_1(\omega_{\leq j}) - j}{S_1(\omega_{\leq j}) - K - 1}$$



Inria
INVENTEURS DU MONDE NUMÉRIQUE

Experimental results

- Description of the model: more than 10^{96} states
- MC probability based on 10^{7} samples: $1.4 * 10^{-6}$
- With N=1000 samples, score function S_2 and $\gamma_0 \approx 1 - \frac{1}{N}$
  - Probability estimator: $1.581 * 10^{-6}$
  - Variance of the estimator: $0.119 * 10^{-6}$
  - Around 100 levels found adaptively.
- Roughly, the number of samples required for IS is between 1000 and 10000 times less important than with MC. => Gain of time

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Ongoing work

- Quantifying performance of importance splitting:
  - Define more complex score functions to improve efficiency
  - Real case studies (biology, robotics?)
- Continuing the development of PLASMA

Inria
INVENTEURS DU MONDE NUMÉRIQUE