Cross entropy optimisation of importance sampling parameters for statistical model checking

Cyrille Jegourel, Axel Legay, Sean Sedwards

CAV 2012, Berkeley

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

Motivation

Objective :

Given a stochastic system, design a procedure for estimating a rare property in a reasonable time with SMC.

Command semantics

Model described as a system of *commands*: (*guard*, *rate*, *action*)

- *guard*: logical predicate over the state
  - enables action
  - applies to a set of states for which the command is enabled
- *rate*: real valued function over set of enabled states
  - rate of exponential dist (CTMC)
  - probability of action (DTMC)
- *action*: update of state
- *state*: assignment of values to variables

PLASMA command language

```
// Repair model based on Example 1 of (Ridder 2005)

ctmc

const int n=3;   // 3 components per type
const double epsilon = 0.1;
const double mu = 1.0;

module type1
state1 : [0..n] init 0;
[] state1 < n -> epsilon*epsilon*(n-state1) : (state1'=state1+1);
[] state1 >= 2 -> mu : (state1'=0);
endmodule

module type2
state2 : [0..n] init 0;
[] state2 < n -> epsilon*(n-state2) : (state2'=state2+1);
[] state2 >= 2 & state1 < 2 -> mu : (state2'=0);
endmodule

...
```

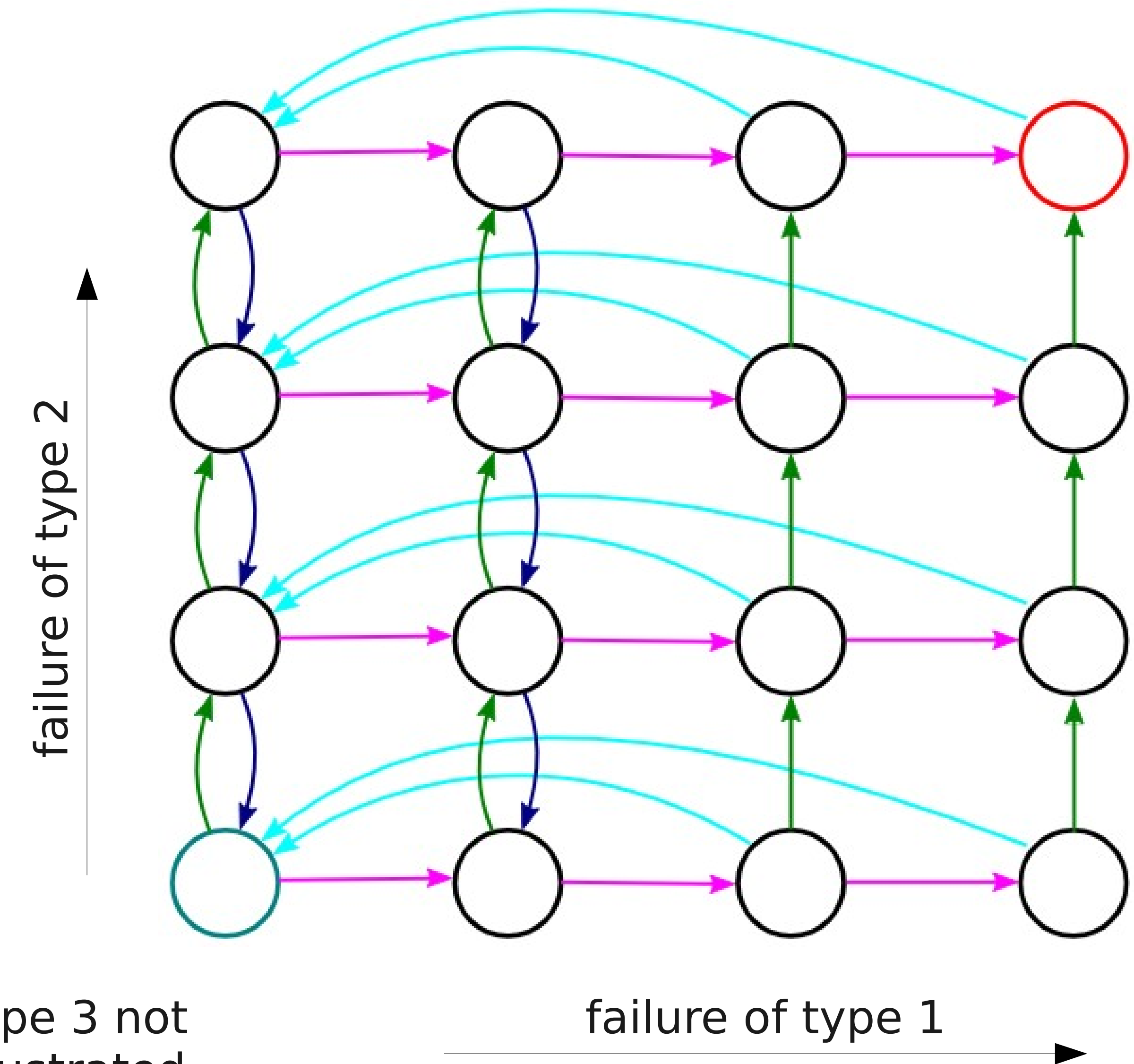
failure type 1

repair type 1

failure type 2

repair type 2

Repair model



Command 1: failure type 1

Command 2: repair type 1

Command 3: failure type 2

Command 4: repair type 2

type 3 not illustrated

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 $z \in \{0,1\}$

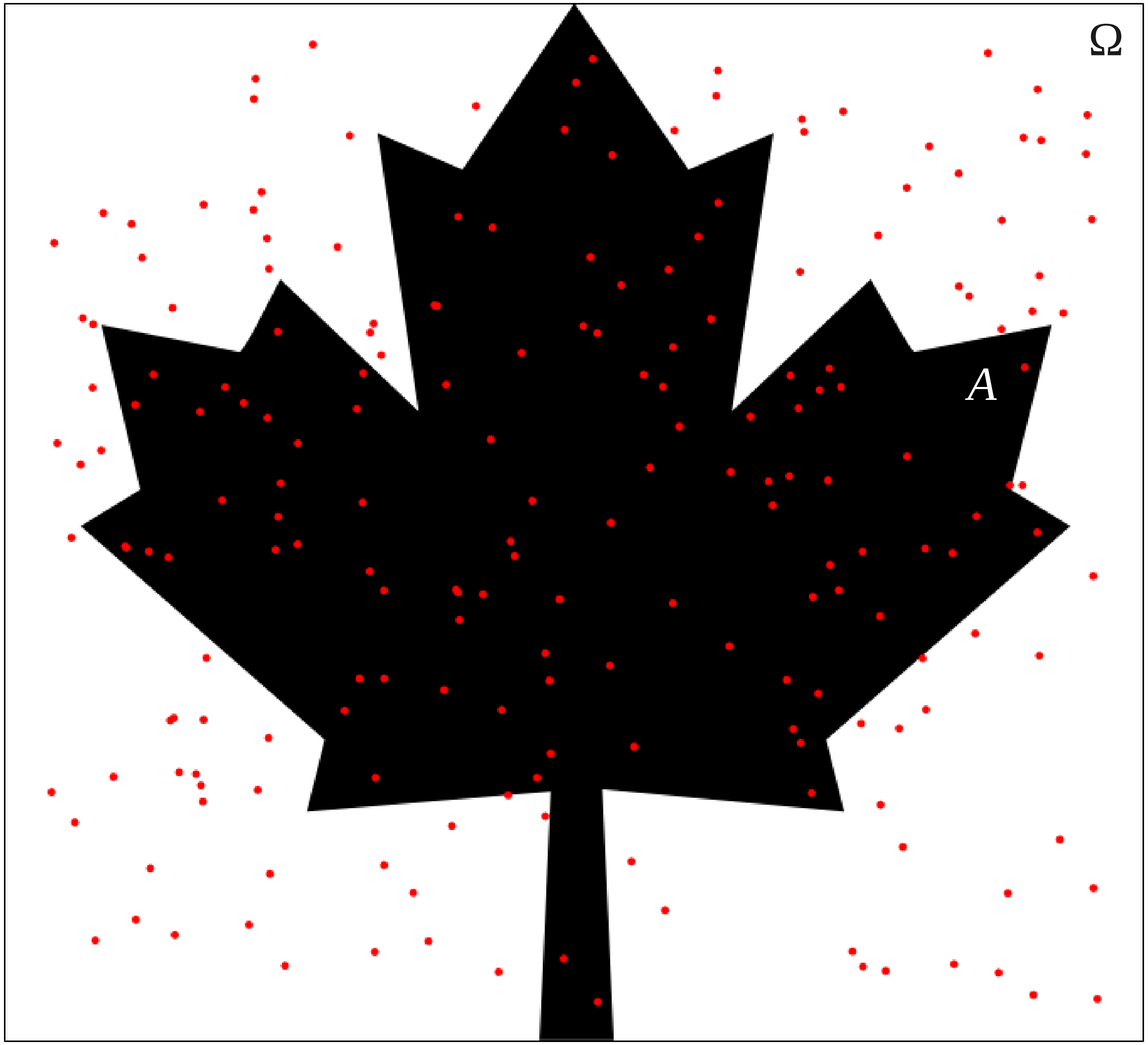
probability measure function

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

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

sample traces generated under *f*

Monte Carlo estimation



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

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

Absolute error = half the size of the confidence interval

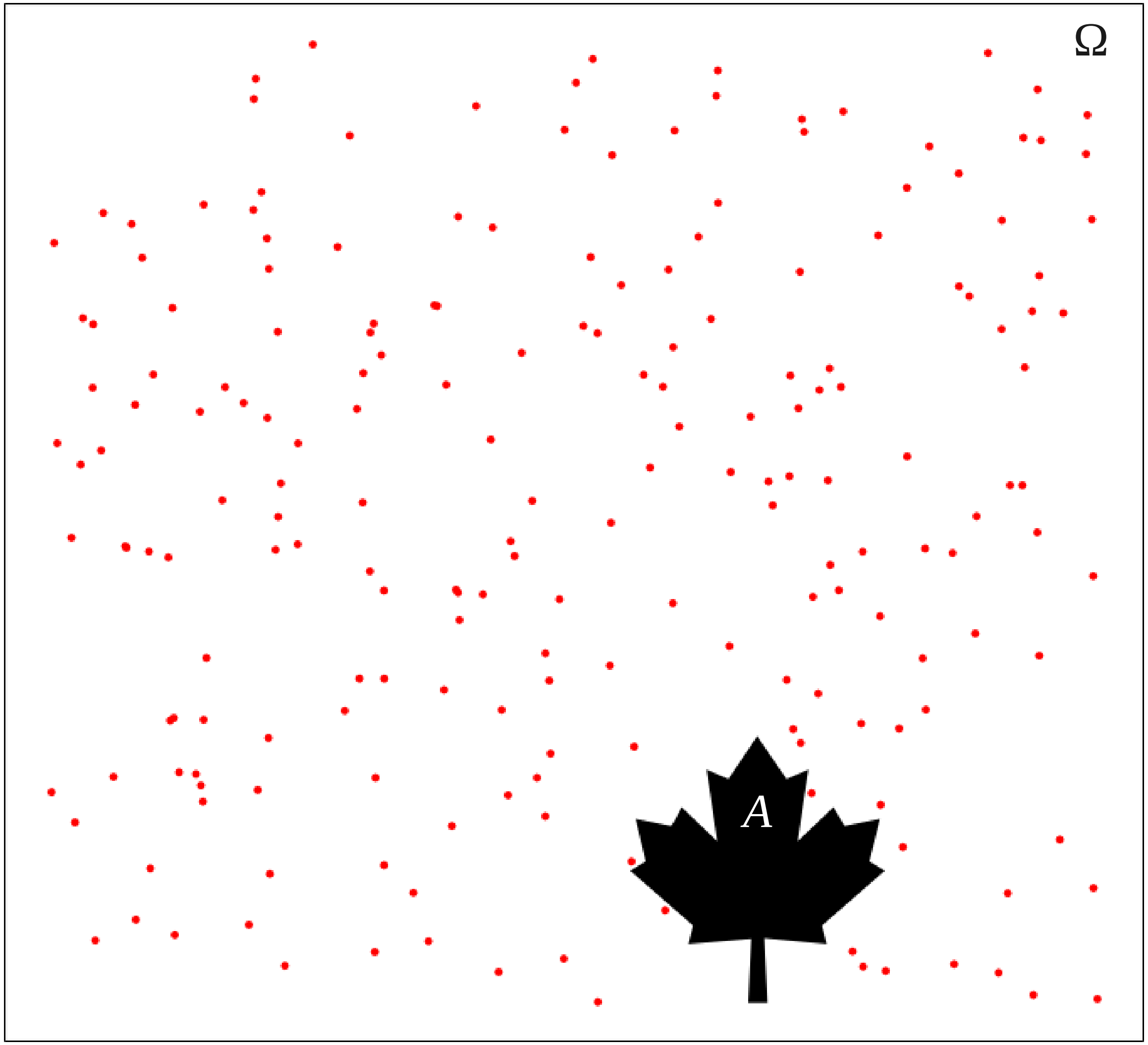
$$\mathrm{AE}\propto\frac{\sqrt{\gamma(1-\gamma)}}{\sqrt{N}}$$

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

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

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*

importance sampling distribution

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

likelihood ratio

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

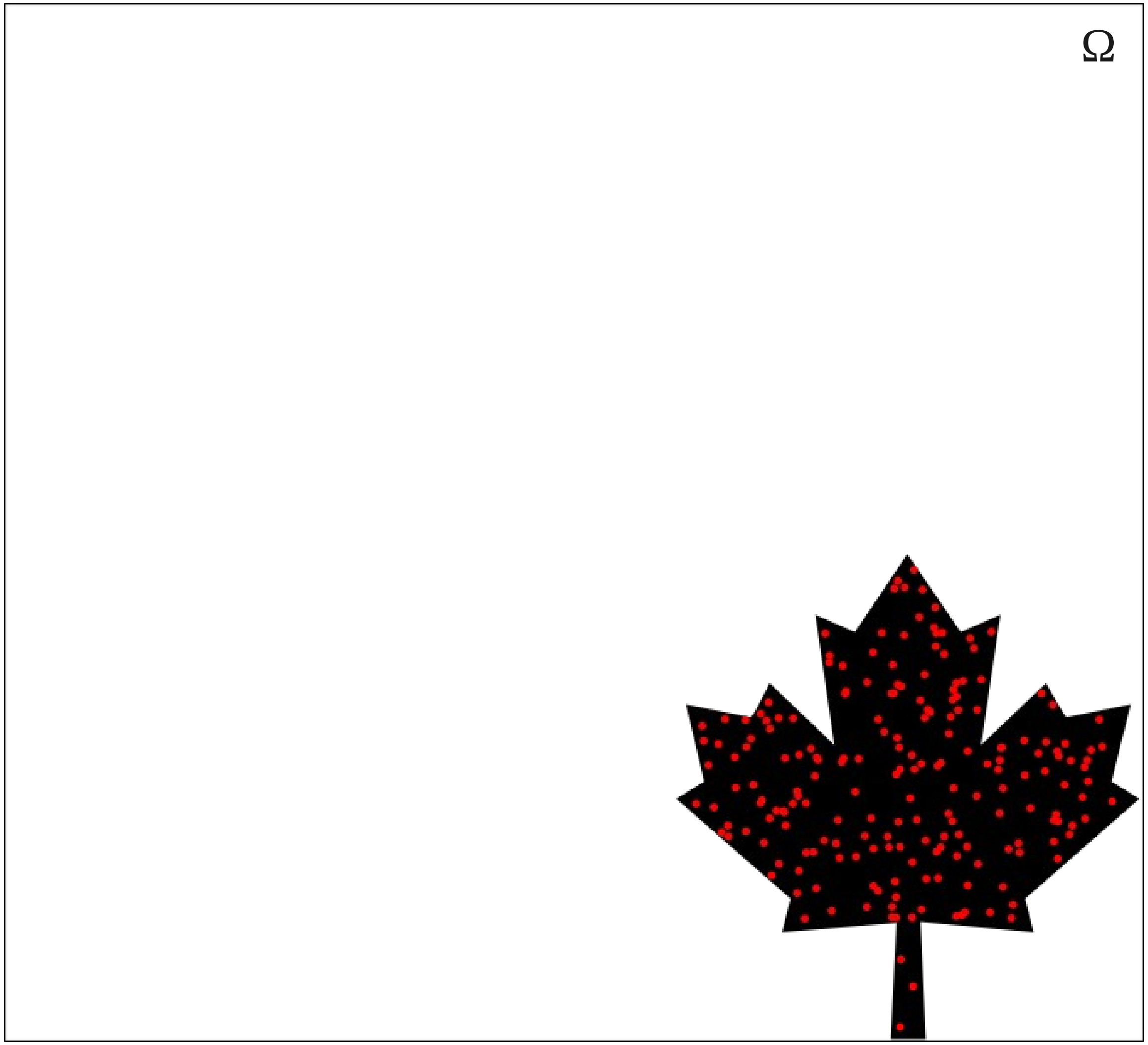
traces generated under *f* '

‘Tilted’ simulation

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

traces generated under f'
(importance sampling dist.)

Optimal importance sampling



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

$$f^{opt}=\frac{zf}{\gamma}$$

f conditioned on the rare event

Summarising...

- IS is a well known technique for reducing the variance of an estimator
- It consists of:
  - modify the dynamics of the system (change of measure)
  - simulate under this new measure
  - unbias the results with the likelihood ratio

- How to perform a good change of measure?

    - Cross-entropy method.

Parametrised models

System of parametrised commands:

$$(guard_i, \lambda_i\, rate_i, action_i) \qquad \lambda_i > 0$$

- each λ controls an action
  - sets of semantically related transitions
  - less precise than individual transitions (algorithm of Ridder) but more tractable
  - parametrisation affects quality of Importance Sampling distribution

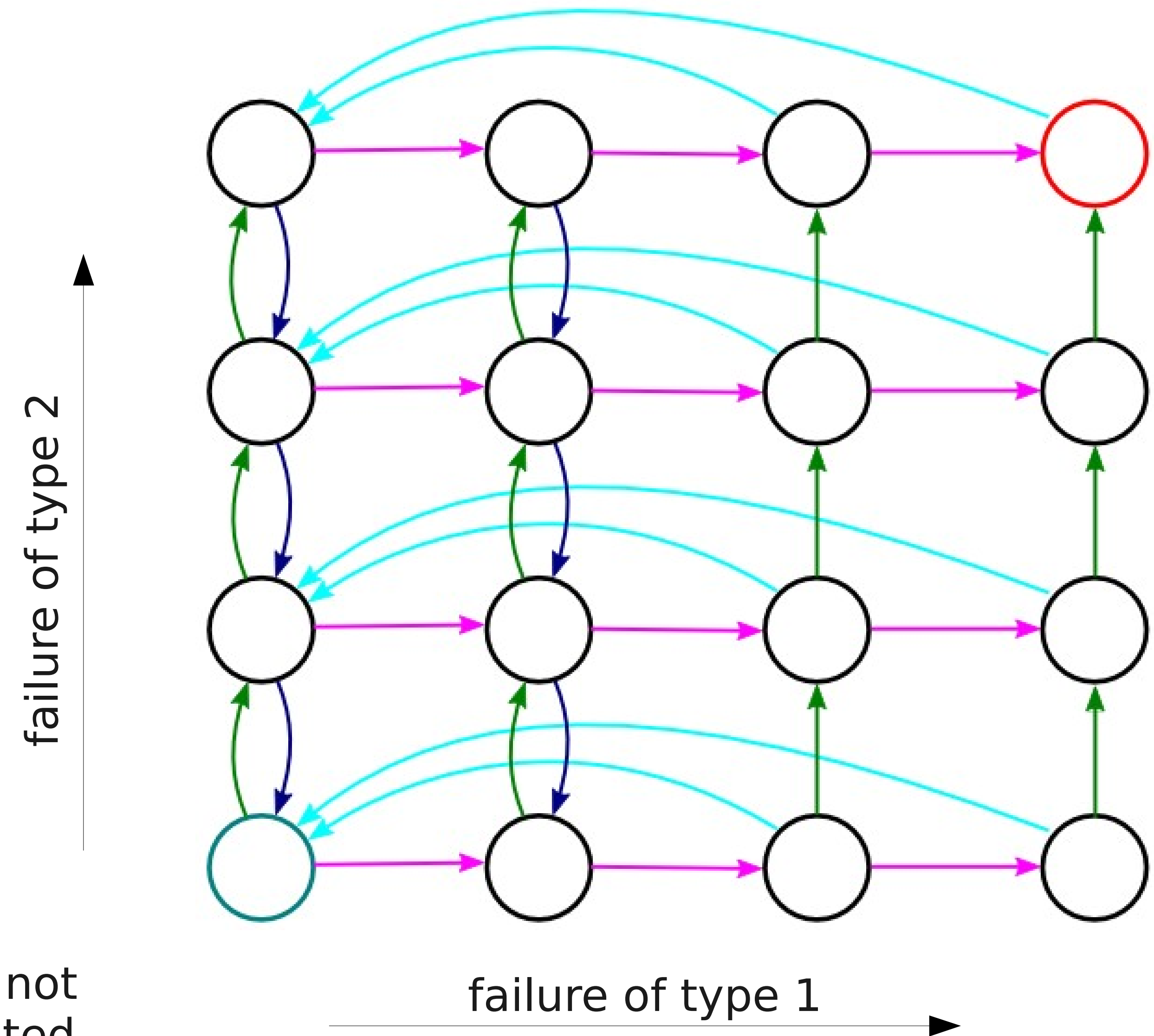
PLASMA parametrised command

$\lambda=[\varepsilon^2, \mu, \varepsilon, \mu]$

$f(x)=[n\text{-}x1, 1, n\text{-}x2, 1]$

```
// Repair model based on Example 1 of (Ridder 2005)

ctmc

const int n=3;   // 3 components per type
const double epsilon = 0.1;
const double mu = 1.0;

module type1
state1 : [0..n] init 0;
[] state1 < n -> epsilon*epsilon*(n-state1) : (state1'=state1+1);
[] state1 >= 2 -> mu*1 : (state1'=0);
endmodule

module type2
state2 : [0..n] init 0;
[] state2 < n -> epsilon*(n-state2) : (state2'=state2+1);
[] state2 >= 2 & state1 < 2 -> mu*1 : (state2'=0);
endmodule

...
```

Parametrised repair model



λ_1 failure type 1

λ_2 repair type 1

λ_3 failure type 2

λ_4 repair type 2

type 3 not illustrated

The Cross-entropy Method (1)

- The Kullback-Leibler divergence:
  - a measure of “distance” between distributions:

$$\mathrm{CE}(g,h) = \mathrm{E}_g[\log \frac{g(\omega)}{h(\omega)}] = \int_{\Omega} g(\omega) \log \frac{g(\omega)}{h(\omega)} d\omega$$

- Goal: system originally parametrised by vector μ,
  - find:

$$\lambda^{opt} \stackrel{\text{def}}{=} \mathrm{argmin}_{\lambda \in S}\, \mathrm{CE}(\frac{z(.)f(.,\mu)}{\gamma}, f(.,\lambda)) \quad (1)$$

  - which is equivalent to find:

$$\lambda^{opt} = \mathrm{argmax}_{\lambda \in S}\, \mathrm{E}_\mu[z(\omega) \log f(\omega,\lambda)] \quad (2)$$

INVENTEURS DU MONDE NUMÉRIQUE

The Cross-entropy Method (2)

- Estimating directly (2) is hard
- Rewrite (2) using Importance Sampling (with L the likelihood ratio):

$$\lambda^{opt} = \text{argmax}_{\lambda \in S}\, \mathrm{E}_{\lambda'}[z(\omega) L(\omega; \mu, \lambda') \log f(\omega, \lambda)] \qquad (3)$$

- Using (3), iteratively construct an estimator:

$$\tilde{\lambda}^{opt} = \lambda^{(j+1)} = \text{argmax}_{\lambda \in S} \sum_{i=1}^{N_j} z(\omega_i^{(j)}) L^{(j)}(\omega_i^{(j)}; \mu, \lambda) \log f(\omega_i^{(j)}, \lambda) \qquad (4)$$

Our algorithm (1)

- Ingredients:
  - a system of *n* guarded commands with state-dependent vector of rate functions: $f(x)=[f_1(x),\dots,f_n(x)]$
  - and corresponding vector of parameters: $\lambda=[\lambda_1,\dots,\lambda_n]$
- In any state *x*, prob of taking command *k*: $\dfrac{\lambda_k f_k(x)}{\langle\lambda,f(x)\rangle}$
- Prob of taking path ω:

number of transitions of type *k* in ω

$$F(\omega,\lambda)=\prod_{k=1}^{n}\left((\lambda_k)^{U_k(\omega)}\prod_{s=1}^{U_k(\omega)}\frac{f_k(x_s)}{\langle\lambda,f(x_s)\rangle}\right)$$

INVENTEURS DU MONDE NUMÉRIQUE

Our algorithm (2)

- **Theorem:** A solution of (5) is almost surely a unique maximum, up to a normalising scalar
  - finding solution is equivalent to solve the convex CE program (4):

$$\frac{dF}{d\lambda_k}(\lambda)=0 \Leftrightarrow \sum_{k=1}^{N} l_i z_i \left(\frac{u_i(k)}{\lambda_k} - \sum_{s=1}^{|\omega_i|} \frac{f_k^i(x_s)}{\langle \lambda, f^i(x_s) \rangle} \right) = 0 \quad (5)$$

With: $l_i = L^j(\omega_i), N^{(j)} = N, z_i = z(\omega_i), u_i(k) = U_k(\omega_i)$

- No closed-form solution, however...

Our algorithm (3)

- Equation (5) leads to the following expression for λ:

$$\forall k \in \{1, \dots, n\}, \qquad \lambda_k = \frac{\sum_{k=1}^{N} l_i z_i u_i(k)}{\sum_{i=1}^{N} l_i z_i \sum_{s=1}^{|\omega_i|} \frac{f_k^i(x_s)}{\langle \lambda, f^i(x_s) \rangle}} \qquad (6)$$

- The right side is still dependent on λ. So,

$$\forall k \in \{1, \dots, n\}, \qquad \lambda_k^{(j+1)} = \frac{\sum_{k=1}^{N} l_i z_i u_i(k)}{\sum_{i=1}^{N} l_i z_i \sum_{s=1}^{|\omega_i|} \frac{f_k^i(x_s)}{\langle \lambda^{(j)}, f^i(x_s) \rangle}} \qquad (7)$$

- Equation (7) has a unique fixed point that is λ^{opt}

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Important details

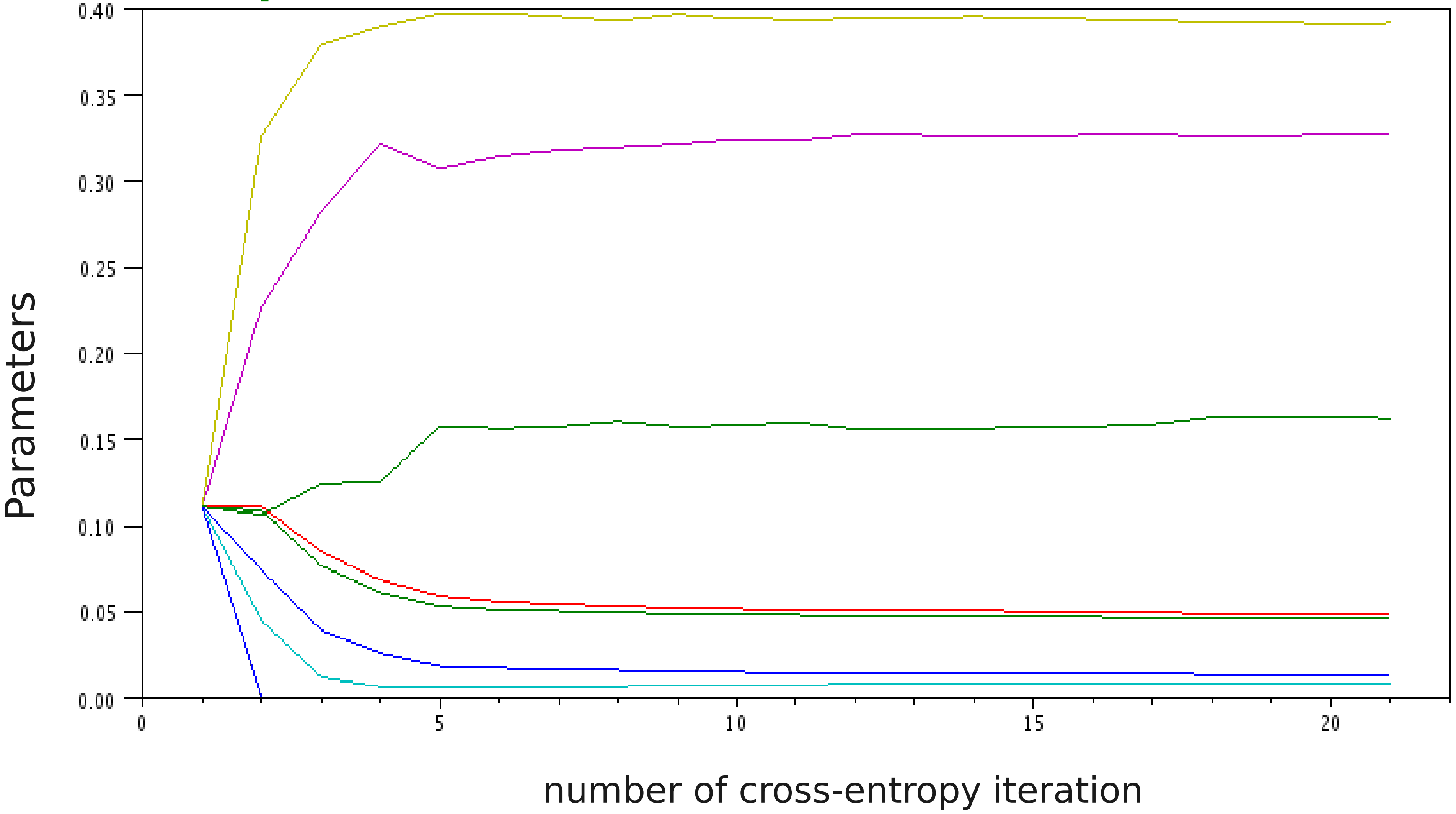
- Initial distribution:
  - the algorithm requires an “adequate start”

  It means that $f(.,\lambda^{(0)})$ must produce at least a few traces satisfying the rare property. Several possibilities, e.g.,
  - equalisation of initial rates
  - random parameters (rare property ≠> rare parameters)
- Smoothing:
  - acts to preserve important but as yet unseen parameters
  - add a small fraction of the initial or previous parameters to every new parameter estimate

INVENTEURS DU MONDE NUMÉRIQUE

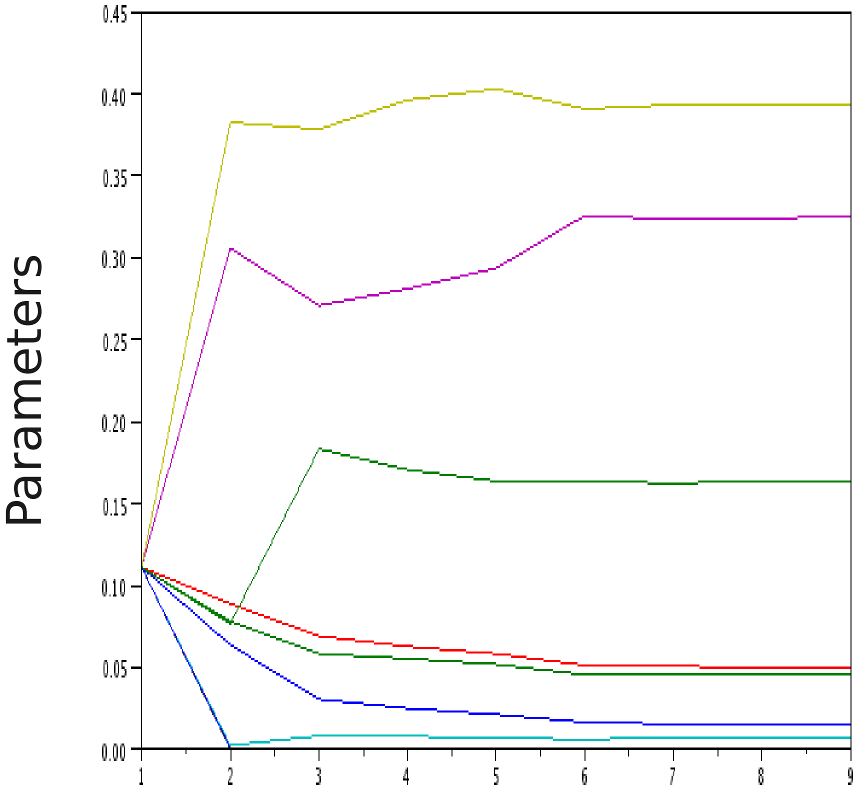
Cross-entropy convergence of parameters

Example 1 of Ridder (2005)

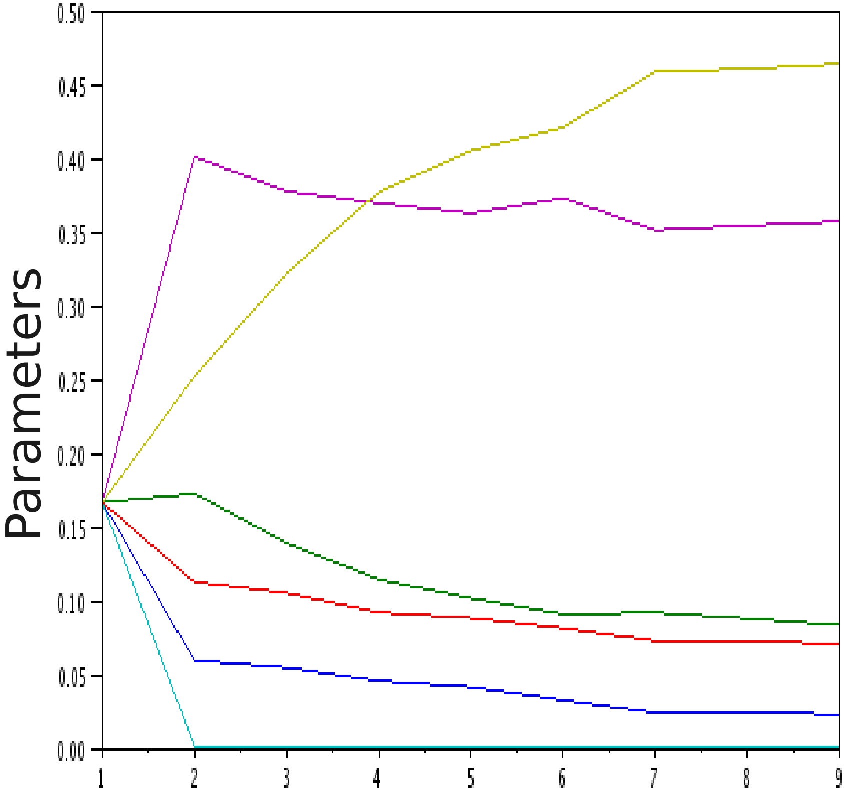


Cross-entropy convergence (2)

- System modeled with 9 commands



- System modeled with 6 commands



number of cross-entropy iteration

Experimental results

- Description of the model: 125 states, 1262 transitions
- Theoretical probability: $1.177 * 10^{-7}$
- Model described by 9 parameters
  - Probability estimator: $1.170 * 10^{-7}$ (+/- $1.0 * 10^{-8}$)
- Model described by 6 parameters
  - Probability estimator: $0.981 * 10^{-7}$ (+/- $2.5 * 10^{-8}$)
- Roughly, the number of samples required for IS is between 1000 and 10000 times less important than with MC. => Gain of time

Ongoing work

- Quantifying performance of importance sampling:
  - Automatise more complex parametrisations to improve efficiency
  - Implement alarms in case of IS failure
  - Real case studies (biology?)
- Continuing the development of PLASMA