Formal Methods in Software Development

Ivano Salvo
and Igor Melatti

Computer Science Department

SAPIENZA
UNIVERSITÀ DI ROMA

Lesson **11**, December 10th, 2019

Probabilistic Systems

Real systems are often dependent on phenomena of a stochastic nature. Here, we address **verification of probabilistic systems**.

By contrast, **probabilistic verification** means no complete coverage ("there is no error with a probability of 90%").

* **Randomized algorithms**: several algorithms (distributed) such leader election use tossing coins to break symmetries.

* Modelling **unreliable or unpredictable** behaviours (ex: message loss, system failures): modelling that with nondeterminism can be too coarse. In late stage of model design, probabilistic valuation can take place of nondet.

* **Performance evaluation**: distribution of inputs, messages, etc. are importat to evaluate quantitative aspects such as waiting time, queue length, expected time between failures.

Verifying Probabilistic Systems

We will see:

- **Markov chains** as generalisation of Kripke structures and in this view we will have a "state based" approach to Markov chains;
- Probabilities in **Linear Time properties** (here probabilities appear at the "semantic level")
- A logic for defining probabilistic properties (here probabilities are in the syntax): **PCTL**.

Quantitative properties: "The probability for delivering a message in the next *t* time units is 98%"

Qualitative properties: A desired event happens almost surely (i.e. with probability 1) or a bad event occurs almost never (i. e. with probability 0): reachability, persistency, repeated reachability.

Lesson 11a:

Markov Chains

Markov Chains: definition

Definition: A (discrete time) **Markov chain** is a tuple $\mathcal{M} = (S, \mathcal{P}, \iota, AP, L)$ where:

S, AP, L as usual are states, atomic propositions and labelling

$\mathcal{P} : S \times S \to [0,1]$ is the **transition probability function**, such that for all $s \in S$, $\sum_{t \in S} \mathcal{P}(s, t) = 1$

$\iota : S \to [0,1]$ is the **initial distribution**, such that $\sum_{s \in S} \iota(s) = 1$

$\mathcal{M}$ is finite if S and AP are finite, and the size of $\mathcal{M}$ is:

$$|\mathcal{M}| = |S| + |\{ (s, t) \in S : \mathcal{P}(s, t) > 0\}|$$

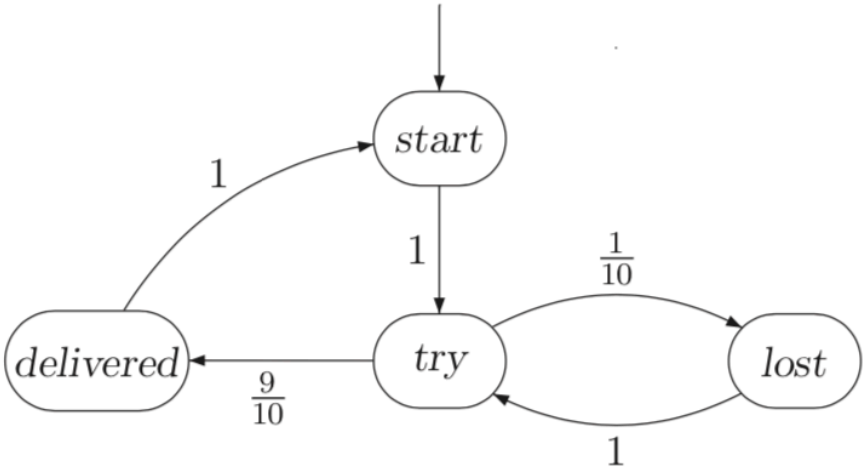
(it is the size of the **underlying digraph**)

We will identify $\mathcal{P}$ with the matrix of probability $[\mathcal{P}(s, t)]_{s,t \in S}$ where the row $\mathcal{P}(s, \cdot)$ contains probability to reach successors of s, and the column $\mathcal{P}(\cdot, s)$ contains probability to enter state s from its predecessors.

States such that $\iota(s)>0$ are **initial states** and it is the probability that system evolution starts in s.

Markov Chains: Example 1

Let us consider an error prone **communication protocol**, that with probability 10% can loose a message. The message is sent until it is eventually delivered.



Probability matrix and initial states (start, try, lost, delivered):

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{10} & \frac{9}{10} \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad \iota_{\text{init}} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Markov Chains: Example 1

Observe that in the underline Kripke structure (without probabilities) we can check LTL or CTL properties, like:

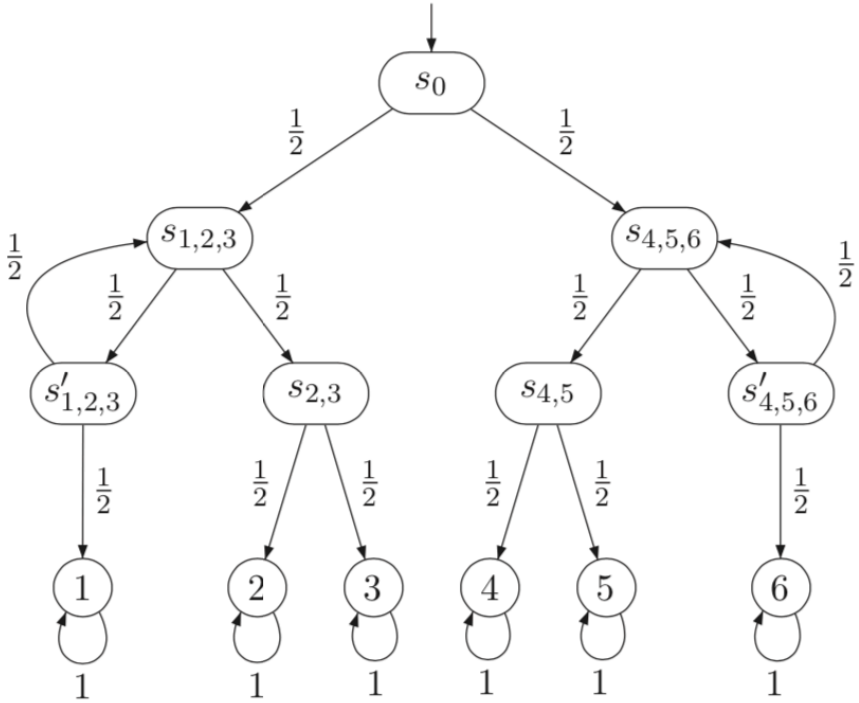
$$\mathbf{G}\,\mathbf{X}^{100}\,\text{delivered} \quad \text{and} \quad \mathbf{E}\,\mathbf{G}\neg\text{delivered}$$

Both these two properties does not hold, even though with very low probability. In particular, the second has probability 0!

Probabilistic model checking allow **quantitative properties** to be checked. **Qualitative** properties are a special case, when we ask for an event to have probability 0 or 1.

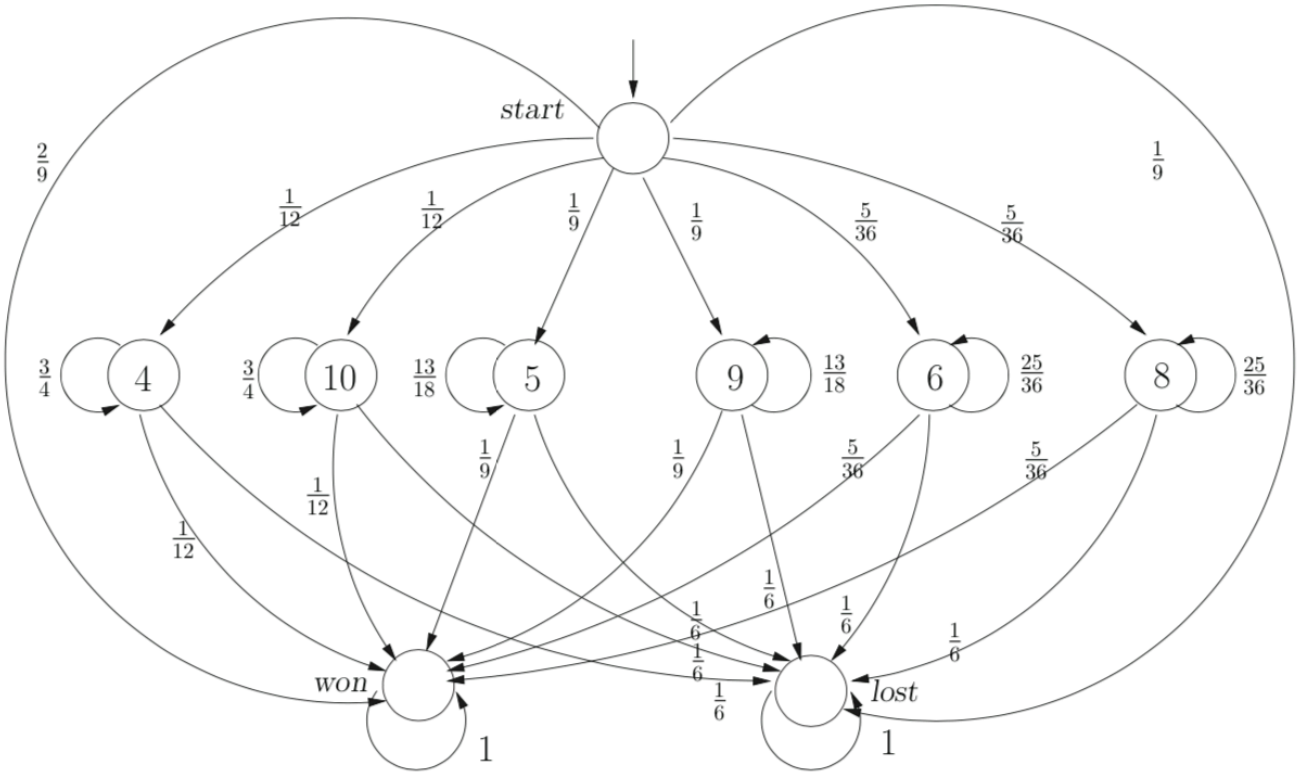
Markov Chains: Example 2

Simulation of a standard **six-sided die** by a fair coin [Knuth&Yao]. Are all outcomes equally likely? Exercise ☺



Markov Chains: Example 3

Simulation of the **Craps Gambling Game**… Player wins on 7, 11 and loses on 2, 3, 12. Otherwise dice are rolled again, until eventually 7 or the point is obtained. 7 player loses, the point wins.



Markov Chains: terminology

$Paths(\mathcal{M})$ denotes the set of paths, $Paths_{fin}(\mathcal{M})$ finite paths. When $\mathcal{M}$ is clear from the context, and s is a state, we can use $Paths(s)$ and $Paths_{fin}(s)$ to denote paths starting at s. In a path $s_0 s_1 \ldots s_n$ we have that

Direct successors of a state s are denoted by $Post(s)$. $Post^*(s)$ is the set of states reachable from s. Similarly, direct predecessors of s are denoted by $Pre(s)$. $Pre^*(s)$ is the set of states backward reachable from s. These notions are naturally extended to sets.

A state s of a MC $\mathcal{M}$ is said **absorbing** if $Post^*(s)=\{s\}$, that is $\mathcal{P}(s, s)=1$ and $\mathcal{P}(s, t)=0$ when $s \neq t$.

A taste of probability: σ-algebras

Definition: A **σ-algebra** is a pair $(O, \mathcal{E})$ where O is a nonempty set (**outcomes**) and $\mathcal{E} \subseteq \mathcal{P}(O)$ is the set of **events** and it contains the emptyset and it is closed under complementation and countable unions. More formally:

- $\varnothing \in \mathcal{E}$,
- If $E \in \mathcal{E}$ then $O \setminus E \in \mathcal{E}$,
- If $E_1, E_2 \ldots \in \mathcal{E}$ then $\cup_{i \geq 1} E_i \in \mathcal{E}$.

Observations: $O \in \mathcal{E}$ as the complement of $\varnothing$. $\mathcal{E}$ is closed under countable intersections, since $\cap_{i \geq 1} E_i = O \setminus \cup_{i \geq 1} (O \setminus E_i)$.

$\mathcal{P}(O)$ is always a σ-algebra and also $\mathcal{E}=\{\varnothing, O\}$.

Definition: A **probability measure** on $(O, \mathcal{E})$ is a function $Pr : \mathcal{E} \to [0,1]$ such that $Pr(O)=1$, and for a family of pairwise disjoint sets: $Pr(\cup_{i \geq 1} E_i) = \sum_{i \geq 1} Pr(E_i)$.

A **probability space** is the triple $(O, \mathcal{E}, Pr)$.

Probability spaces: properties

Example: Let us consider the experiment of tossing a fair coin once. The set O of outcomes is {head, tail}. We can consider the singletons {head}, {tail} as the set of events. The smallest σ-algebra containg such events is $\mathcal{P}(\{\text{head, tail}\})$ with

$Pr(\varnothing)=0$, $Pr(\{\text{head}\})=Pr(\{\text{tail}\})=1/2$, and $Pr(\{\text{head,tail}\})=1$. □

When O is countable, then fixing a function $\mu : O \to [0,1]$, such that $\sum_{e\in O} \mu(e) = 1$ defines a probability measure on $(O, \mathcal{P}(O))$, defined by $Pr(E)= \sum_{e\in E} \mu(e)$.

Since $E \cup (O \setminus E)=O$ and they are two disjoint sets, $Pr(O \setminus E)=1-Pr(E)$. In particular, $Pr(\varnothing)=1 - Pr(O)=0$.

Probability measures are **monotonic**: If $E \subseteq F$, then $Pr(F)=Pr(E)+Pr(F \setminus E) \geq Pr(E)$.

For each set $P \subseteq \mathcal{P}(O)$, there exists a **smallest σ-algebra $\mathscr{E}_P$** that contains P. $\mathscr{E}_P$ is called the σ-algebra **generated** by P, and P is the **basis**.

σ-algebras and Markov chains

Definition: The **cylinder set** of a finite path $\pi = s_0 s_1 \ldots s_n$ is $Cyl(\pi) = \{\pi' \mid \pi' = \pi\pi''\}$.

The **σ-algebra $\mathcal{E}_{\mathcal{M}}$ associated with a Markov chain $\mathcal{M}$** is generated by all $Cyl(\pi)$, for π finite path in $\mathcal{M}$.

$$Pr(Cyl(s_0 s_1 \ldots s_n)) = \iota(s_0) \prod_{0 \le i < n} \mathcal{P}(s_i, s_{i+1})$$

Notation: We will use LTL-like syntax to denote events in the probability space $(\text{Path}_{\mathcal{M}}, \mathcal{E}_{\mathcal{M}}, Pr)$.

For example, if $B \subseteq S$, "**F** B" is the set of paths that reach the set B after a finite number of steps and "**GF** B" is the event of visiting B infinitely often. Sometimes we will write $\pi \models \varphi$ for $\pi \in \varphi$ and we denote with $Pr(s \models \varphi)$ the probability of φ to hold in the state s, that is $Pr(\{\pi \in \text{Path}(s) \mid \pi \models \varphi\})$.

Reachability problems

As for classical Model Checking, one of the basic problems is reachability: here, the problem is to compute the probability of reaching a given set of states $B \subseteq S$.

$\mathrm{Path}(\mathbf{F}\ B) = \mathrm{Path}_{fin}(\mathcal{M}) \cap (S \setminus B)^* B$ is the set of path that reach B.

$$Pr(\mathbf{F}\ B) = \sum_{\pi \in \mathrm{Path}(\mathbf{F}\ B)} Cyl(\pi)$$

Example: Let us compute the probability of reaching the state delivered in the simple communication protocol: the path has the form:

$$\pi = \text{start try (lost try)}^n \text{ delivered}$$

from which we derive:

$$Pr(\mathbf{F}\ \text{delivered}) = \sum_{n \geq 0} (1/10)^n\, 9/10 = 1$$

that is quite intuitive: any message will be eventually delivered. If we put a bound on retransmissions, say 3, we have:

$$Pr(\mathbf{F}\ \text{delivered}) = 9/10 + 1/10 * 9/10 + 1/100 * 9/10 = 0.999$$

Computing probabilities

Lex $x_s = Pr(s \models \mathbf{F}\, B)$. For s∈B, $x_s = 1$. For $s \in S \setminus B$, we have:

$$x_s = \sum_{t \in S \setminus B} P(s, t) \cdot x_t + \sum_{u \in B} P(s, u) \qquad (*)$$

This is a sort of "probabilistic expansion law". By considering only states in $S' = Pre^*(B) \setminus B$, $(*)$ $x = (x_s)_{s \in S'}$ becomes: $x = \mathbf{A}\, x + \mathbf{b}$, where $\mathbf{A}$ is $(\mathcal{P}(s, t))_{s,t \in S'}$ and $\mathbf{b}$ is the probability of reaching S' in one step which can be rewritten as $(\mathbf{I} - \mathbf{A})\, x = \mathbf{b}$, where $\mathbf{I}$ is the identity matrix of size $|S'| \times |S'|$.

Example [Communication Protocol]: let B = {delivered} and S'={start, try, lost}. We can easily obtain the following equations:

$$x_{\text{start}} = x_{\text{try}} \qquad x_{\text{try}} = 1/10\, x_{\text{lost}} + 9/10 \qquad x_{\text{lost}} = x_{\text{try}}$$

that correspond to the system (the solution is 1 for all states):

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -\frac{1}{10} \\ 0 & -1 & 1 \end{pmatrix} x = \begin{pmatrix} 0 \\ \frac{9}{10} \\ 0 \end{pmatrix}$$

Algorithm

First compute the set S'. This can be done simply by a backward visit starting from B.

Then generate the matrix **A** and the vector **b** and solve the linear system $(\mathbf{I} - \mathbf{A})x = \mathbf{b}$. This system can have more than one solutions if **I** – **A** is singular.

The solution we are interested is the **least solution** in [0,1].

To avoid this problem, we considered an **iterative method** (instead of direct methods) for a more general problem **constrained reachability**, that is for until properties

Iterat. Constrained Reachability

Let $B, C \subseteq S$. We consider the problem of reaching B via a finite path fragment in C, that is $C \mathbf{U} B$. For $n \geq 0$, the event $C \mathbf{U}^{\leq n} B$ is the same, but it is required that B is reached in at most n steps.

We partition S as follows:

- $S \setminus (C \mathbf{U} B) \subseteq S_0 \subseteq \{ s \in S \mid Pr(s \models C \mathbf{U} B)=0 \}$
- $B \subseteq S_1 \subseteq \{ s \in S \mid Pr(s \models C \mathbf{U} B)=1 \}$
- $S_? = S \setminus (S_0 \cup S_1)$

Theorem: The vector $(x_s)_{s \in S?}$ is the least fixed point of the operator $\boldsymbol{\Upsilon} : [0,1]^n \rightarrow [0,1]^n$ defined by: $\boldsymbol{\Upsilon}(y) = \mathbf{A}y + \mathbf{b}$, where n is the cardinality of $S_?$, $\mathbf{A}$ is the probability transition restricted on states in $S_?$, and $\mathbf{b}$ is the vector of probability of enter B in one step. Furthermore, if x^0 is $\mathbf{0}$ and $x^{n+1} = \boldsymbol{\Upsilon}(x^n)$, we have:

- $x_s^n = Pr(s \models C \mathbf{U}^{\leq n} S_1)$,
- $x_s^0 \leq x_s^1 \leq x_s^2 \leq \ldots$,
- $x = lim_{n \rightarrow \infty} x^n$

Iterative Algorithm

The previous theorem suggests an iterative algorithm to compute x_s. $x^0 = \mathbf{0}$ and $x^{n+1} = \boldsymbol{\Upsilon}(x^n)$. Since this sequence converges, we can stop when $| x^{n+1} - x^n | < \varepsilon$, for some small tolerance ε.

Remark: Sets S_0 and S_1 are **not uniquely** identified. For example, $S_0 = S \setminus (C \mathbf{U} B)$ and $S_1 = B$ suffices. However, the largest S_0 and S_1, the faster is the convergence (smaller matrices, etc.) A reasonable choice is:

$S_0 = \{ s \in S \mid Pr(s \vDash C \mathbf{U} B)=0 \}$ and $S_1 = \{ s \in S \mid Pr(s \vDash C \mathbf{U} B)=1 \}$.

Bounded Until Properties. Taking $S_0 = S \setminus C \mathbf{U} B$ and $S_1 = B$ and $S_? = C \setminus B$ we have that $x^n(s) = \Pr(s \vDash C \mathbf{U}^{\leq n} B)$.

Remark: The n^{th} power of $\mathbf{A}$ contains probabilities to reach a state in exactly n steps. More precisely, $\mathbf{A}^n(s, t)$ is the sum of probabilities of all paths of the form $s=s_0 s_1 \ldots s_n=t$.

In other words: $\mathbf{A}^n(s, t)=\Pr(s \vDash S \mathbf{U}^{=n} t)$

Transient probabilities

If B=$\varnothing$, S_0=S_1= $\varnothing$ and C = $S_?$= S then $\mathbf{A}$=$\mathcal{P}$, then $\mathcal{P}^n(s, t)$ is the probability of being in the state t after n steps starting in s, that is $Pr(s \models S\ \mathbf{U}^{=n}\ t)$.

The probability of $\mathcal{M}$ of being in state t after n steps is the **transiet probability of state** t, defined by $\theta_n(t)=\sum_{s \in S} \mathcal{P}^n(s, t)\ \iota(s)$ and $\theta_n=\mathcal{P}^n \cdot \iota$.

θ_n can be used to compute $\mathbf{F}^{\leq n}\ B$ and $C\ \mathbf{U}^{\leq n}\ B$.

Let us consider $\mathcal{M}'$ where we substitute all outgoing arrows from $b \in B$ with self loops of probability 1.

Let us consider $\mathcal{M}''$ where we substitute in $\mathcal{M}'$ all arrows exiting from $c \in S \setminus (C\ \mathbf{U}\ B)$ with self-loop of probability 1.

We have: $Pr_{\mathcal{M}}(s \models \mathbf{F}^{\leq n}\ B) = Pr_{\mathcal{M}'}(s \models \mathbf{F}^{=n}\ B)$ and $Pr_{\mathcal{M}}(s \models C\ \mathbf{U}^{\leq n}\ B) = Pr_{\mathcal{M}''}(s \models C\ \mathbf{U}^{=n}\ B)$.

Lesson 11b:

Qualitative properties

Qualitative properties

Qualitative properties require some event to happen with probability 1 or, dually, check if some event occurs with probability 0.

Most of qualitative properties can be established just looking at the underlying digraph, because in a ***finite*** Markov chain ***almost surely*** paths eventually enter in a Bottom Strongly Connected Component (BSCC).

Persistence Properties. The event **GF** B is measurable. This event can be written as a countable intersections of countable unions of cylinder sets (prove this equality is an easy ex):

$$\mathbf{GF}\ B = \cap_{n \geq 0} \cup_{m \geq n} Cyl(\text{“}m{+}1^{th}\text{ state is in B”})$$

Persistence properties of the form **FG** B are measurable as the complement of **GF** B. As a matter of fact, **FG** $B = S \setminus (\mathbf{GF}\ S \setminus B)$.

Probabilistic Choice & Fairness

In a Markov chain, if a state t is visited infinitely often, then almost surely all finite path fragments starting in t will be taken infinitely often. Here "almost surely" has to be read as conditional probability: an event E **holds almost surely under another event** D, if $Pr(D) = Pr(E \cap D)$.

Theorem: Let $\mathcal{M}$ a MC, and $s,t \in S$. Then:

$$Pr(s \models \mathbf{GF}\ t) = Pr(\bigwedge_{\pi \in \mathrm{FinPath}(t)} \mathbf{GF}\ \pi)$$

In particular, the above theorem says that **each transition** (t, t') such that $\mathcal{P}(t, t')>0$ will be taken almost surely. In this sense, **probabilistic choice is strongly fair**.

Theorem: Let $\mathcal{M}$ be a MC, and $s \in S$. Then:

$$Pr(\{\pi \in \mathrm{Path}(s) \mid inf(\pi) \in \mathrm{BSCC}(\mathcal{M})\} = 1$$

In every MC, almost surely, a path ends in a BSCC of $\mathcal{M}$.

Almost sure reachability

The problem of **almost sure reachability** amounts to determine the set of states that reach a given set of goal states B almost surely.

Theorem: Let $\mathcal{M}$ be a finite MC, $s \in S$, and $B \subseteq S$. Then the following statements are equivalent:

- $Pr(s \models \mathbf{F}\ B) = 1$
- $Post^*(t) \cap B \neq \varnothing$ for each $t \in Post^*(s)$
- $s \in S \setminus Pre^*(S \setminus Pre^*(B))$

This theorem gives a purely graph-theoretic characterisation of almost-sure reachability. Observe that from s such that $Pr(s \models \mathbf{F}\ B) = 1$ we cannot go outside $Pre^*(B)$.

Algorithm: Build the MC $\mathcal{M}_B$ where all states in B are made absorbing. Then use two backward reachability on $\mathcal{M}_B$ to compute the set of states $S \setminus Pre^*(S \setminus Pre^*(B))$ [the first from B and the second from $S \setminus Pre^*(B)$]

Qualit. constrained reachability

The problem of **qualitative constrained reachability** amounts to determine the sets of states S_0 and S_1 such that:
$S_0 = \{ s \in S \mid Pr(s \models C \mathbf{U} B)=0 \}$ and $S_1 = \{ s \in S \mid Pr(s \models C \mathbf{U} B)=1\}$.

S_0 corresponds to the set of states satisfying $\neg \mathbf{E}\ (C \mathbf{U} B)$ and can be computed by a backward reachability from B.

As for S_1, we reduce the problem to an almost sure reachablity in a slightly modified Markov chain $\mathcal{M}'$. We make absorbing all states in B and in $S \setminus (C \mathbf{U} B)$.

- $\Pr_{\mathcal{M}}(s \models C \mathbf{U} B) = \Pr_{\mathcal{M}'}(s \models \mathbf{F} B)$ for all $s \in C \setminus B$
- $\Pr_{\mathcal{M}}(s \models C \mathbf{U} B) = \Pr_{\mathcal{M}'}(s \models \mathbf{F} B) = 1$ for all $s \in B$
- $\Pr_{\mathcal{M}}(s \models C \mathbf{U} B) = \Pr_{\mathcal{M}'}(s \models \mathbf{F} B) = 0$ for all $s \in S \setminus (C \setminus B)$

This give a polynomial algorithm (the transformation from $\mathcal{M}$ to $\mathcal{M}'$ is clearly linear in the size of $\mathcal{M}$).

Qualitative repeated reachability

Corollary: Let $\mathcal{M}$ be a finite MC, $s \in S$, and $B \subseteq S$. Then the following are equivalent:

- $Pr(s \models \mathbf{G}\,\mathbf{F}\,B) = 1$
- $T \cap B \neq \varnothing$ for each BSCC T reachable from s.
- $s \models \mathbf{AG}\,\mathbf{EF}\,B$.

Corollary: Let $\mathcal{M}$ be a finite MC, $s \in S$, and $B \subseteq S$ and let U be the union of all BSCC T of $\mathcal{M}$ such that $T \cap B \neq \varnothing$. Then:

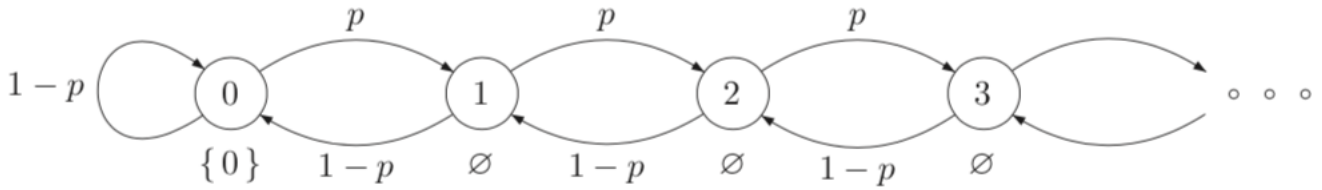
$$Pr(s \models \mathbf{G}\,\mathbf{F}\,B) = Pr(s \models \mathbf{F}\,U)$$

Counterexample: infinite MC

In infinite Markov chains, the probability of visit a state infinitely often could be always 0. We can have that $Pr(s \models \mathbf{F}\ B) > 0$ and $Pr(s \models \mathbf{GF}\ B) = 0$.

Let us consider the one-dimensional random walk below, where $p \in]0,1[$ and $\mathcal{P}(n, n+1) = p$, $\mathcal{P}(n, n-1) = 1 - p$ $(n>0)$, and $\mathcal{P}(0, 0)=1 - p$.

For $p \leq 1/2$, we have that $Pr(n \models \mathbf{F}\ 0) = Pr(n \models \mathbf{GF}\ 0) = 1$, whereas if $p>1/2$ it is more likely to move to the right than to the left and one can show that $Pr(n \models \mathbf{F}\ 0) < 1$ and $Pr(n \models \mathbf{GF}\ 0) = 0$.



Lesson 10

That's all Folks…

…Questions?