An Introduction to Probablistic Automata

Marielle Stoelinga

Fabian Klopfer

Probabilistic Modelling Seminar

November 19, 2019

TOC

Intro

"Probablistic Automata (PAs) provide a mathematical framework for the specification and analysis of asynchronous, concurrent systems with discrete probabilistic choice."

- Is the probability that an error occurs small enough?
- What is the minimal probability that the system responds within 3 seconds?

Example

- E.g. randomized distributed dining philosophers [1]

```
while true
  do think;
    do trying:=true or die od;
    while trying
      do draw a random element s of {Right,Left};
           ***with equal probabilities ***
         wait until s chopstick is down
           and then lift it;
         if R(s) chopstick is down
           then lift it;
             trying:=false
           else
             put down s chopstick
         fi
      od;
    eat;
    put down both chopsticks
      *** one at a time, in an arbitrary order ***
  od.
```

- Bayesian integration in sensorimotor learning [2]
- Hierarchical Bayesian inference in the Visual cortex [3]

A word on non-determinism

Non-Determinism can be divided into internal and external:

Internal choices of the automaton independent of its environment

External choices of the automaton influenced by its environment

Non-Probabilistic Automata

A Non-Propabilistic Automaton (NA) consists of a 4-Tuple $A = (S, S_0, \text{sig}, \Delta)$ with

- S the set of states
- S_0 the initial states
- $\text{sig} = (V, I)$
  where $V = \{\text{external/visible actions}\}$ and $I = \{\text{internal actions}\}$.
  Define Actions $A = V \cup I$
- a transition relation $\Delta \subseteq S \times A \times S$ and for $a \in A, s, s' \in S$ write $s \xrightarrow{a} s'$

Example

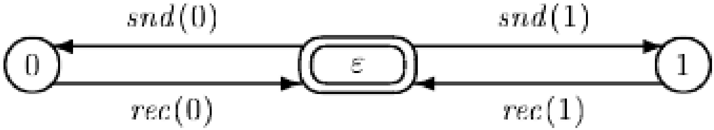


Figure 1: A channel automaton

Probability distribution

A Probability distribution is a function

$$\mu : x \to [0, 1]$$

with

$$\sum_i \mu(x_i) = 1 \qquad x_i \in X$$

The support of μ is the set

$$\{x \in X | \mu(x) > 0\}$$

with probability greater than $0.\bar{0}$. Write

$$\mathrm{D}(X) = \{x_1 \mapsto \mu(x_1), \ldots\}$$

leaving out elements with probability $0.\bar{0}$, i.e. only the support.

PA Definition

A Probablistic Automa (PA) consists of a 4-Tuple $PA = (S, S_0, \text{sig}, \Delta)$ with

- S the set of states
- S_0 the initial states
- $\text{sig} = (V, I)$
  where $V = \{\text{external/visible actions}\}$ and $I = \{\text{internal actions}\}$.
  Define Actions $A = V \cup I$
- a transition relation $\Delta \subseteq S \times A \times D(S)$ and for $a \in A, s \in S, \mu \in D(S)$ write $s \xrightarrow{a} \mu$ for $(s, a, \mu) \in \Delta$

Example

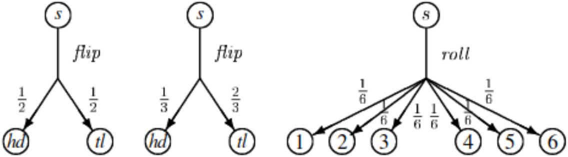


Figure 2: Transitions representing a fair coin flip, an unfair coin flip and a fair dice roll.

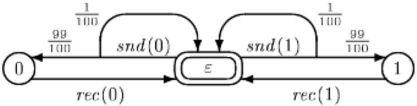


Figure 3: A lossy channel PA

Probabilistic vs. Non-Deterministic choice

- Probabilistic choices are specified within, non-deterministic choices are specified between transitions
- Non-Determinism is chosen when we decide not to specify how a choice is made (e.g. when the environment dictates the transition)
- Probabilities are chosen when all laws of probability theory are satisfied including the Kologmorov axioms:
  - $\forall x \in X : P(x) \in \mathbb{R}, P(x) \geq 0, P(x) \leq 1$
  - $\sum_i P(x_i) = 1$
  - $P\left(\bigcup_{i=1}^{\infty} x_i\right) = \sum_{i=1}^{\infty} P(x_i)$

  and the law of large numbers with arithmetic mean $\bar{x} = \frac{1}{n}\sum_i^n (X_i - E[X_i])$:

$$P\left(\lim_{n\to\infty} \bar{x}_n = \mu\right) = 1.$$

Probabilistic and Non-Deterministic choice are orthogonal

Example

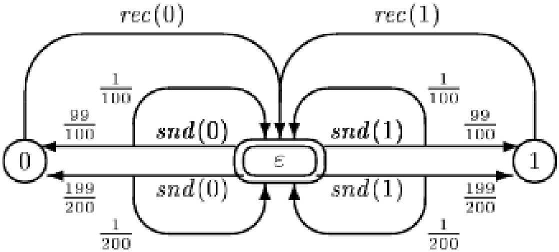


Figure 5: A lossy channel PA with partially unknown probabilities

Timing & the PTA

A Probablistic Timed Automata (PTA) is a PA with time passage actions. Split the internal actions I into discrete actions A_D and time-passage actions A_T with the following requirements to the PTA with $s, s', s'' \in Smd, d' \in \mathbb{R}_{>\not{0}} : ' < :$

- each transition labeled with a time-passage action leads to a distribution that satisfies the Kologmorov axioms
- (Time determinism) if $s \xrightarrow{d} s' \wedge s \xrightarrow{d} s'' \Leftarrow s' = s''$
- (Wang's axiom) $s \xrightarrow{d} s'' \Leftrightarrow \exists s' : s \xrightarrow{d'} s' \wedge s' \xrightarrow{d-d'} s''$

Example

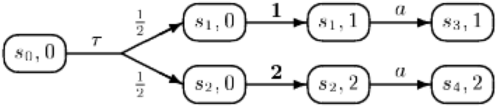


Figure 6: A part of a PTA

$$
\begin{aligned}
&(s_0, 0) \xrightarrow{\tau} \{(s_1, 0) \mapsto \tfrac{1}{2}, (s_2, 0) \mapsto \tfrac{1}{2}\}, \\
&(s_1, d) \xrightarrow{d'} (s_1, d + d'), && \text{if } d + d' \leq 1, \\
&(s_2, d) \xrightarrow{d'} (s_2, d + d'), && \text{if } d + d' \leq 2, \\
&(s_1, 1) \xrightarrow{a} (s_3, 1), \\
&(s_2, 2) \xrightarrow{a} (s_4, 1).
\end{aligned}
$$

Parallel Composition

The Parallel composition operator $||$ allows to construct a PA from several component PA.

Define joint probability distribution with μ as D(X) and φ as D(Y) as:

$$\mu \times \varphi := D(X \times Y) \text{ with } (\mu \times \varphi)(x, y) = \mu(x) \cdot \varphi(y)$$

Two PAs are compatible if $I_A \cap A_B = \emptyset = A_A \cap I_B$

Parallel Composition

Define parallel composition of two *compatible* PAs A||B by:

- $S_{A||B} = S_A \times S_B$
- $S^0_{A||B} = S^0_A \times S^0_B$
- $\mathrm{sig}_{A||B} = (V_A \cup V_B, I_A \cup I_B)$
- $\Delta_{A||B}$ as the set of transitions $(s_1, s_2) \xrightarrow{a} \mu_1 \times \mu_2$ such that at least one of the following requirements is met:
  - $a \in V_a \cap V_B \wedge s_1 \xrightarrow{a} \mu_1 \in \Delta_A \wedge s_1 \xrightarrow{a} \mu_2 \in \Delta_B$
  - $(a \in A_A \setminus A_B \vee a \in I_A) \wedge s_1 \xrightarrow{a} \mu_1 \in \Delta_A \wedge \mu_2 = \{s_2 \mapsto 1\}$
  - $(a \in A_B \setminus A_A \vee a \in I_B) \wedge s_2 \xrightarrow{a} \mu_2 \in \Delta_B \wedge \mu_1 = \{s_1 \mapsto 1\}$

Example I

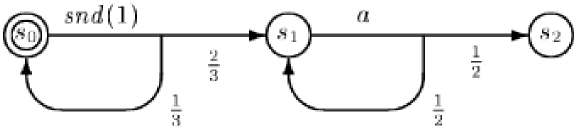


Figure 7: A sender PA

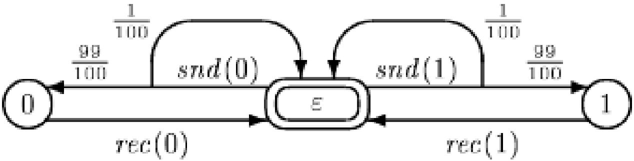


Figure 3: A lossy channel PA

Example II

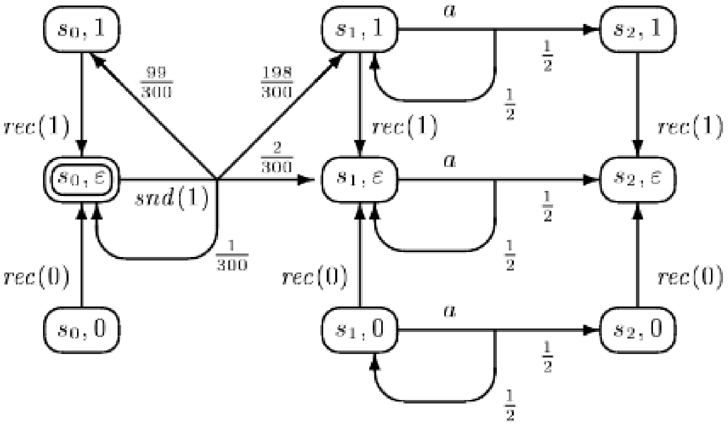


Figure 8: Parallel composition

Basic Steps

1. resolve non-deterministic choices probabilistically using adversaries.
2. abstract non-visible elements by removing the states and internal actions

An adversary is basically a cycle free discrete-time markov chain.

NA: Path & Trace

A Path of an NA A is a sequence

$$\pi = s_0 a_1 s_1 a_2 \ldots$$

where $s_0 \in S_A^0, s_i \in S_A, a_i \in A_A$ and $s_i \xrightarrow{a_{i+1}} s_{i+1} \in \Delta_A$

A trace is a sequence of external actions obtained from a path by omitting the states and internal actions. Denote the set of traces of A by *trace(A)*

PA: Probabilistic Path and Trace

A probabilistic path (or just Path) of a PA B is a sequence

$$\pi = s_0 a_1 \mu_1 s_1 a_2 \mu_2 \dots$$

where $s_0 \in S_B^0, s_i \in S_B, a_i \in A_B$ and $s_i \xrightarrow{a_{i+1}} \mu_{i+1} \in \Delta_B$ and $\mu(s_{i+1}) > 0$.
Let $\text{last}(\pi)$ denote the last state of a finite path,
$|\pi| \in \mathbb{N} \cup \infty$ the number of actions occurring in π,
$\text{Path}^*(A)$ the set of finite paths pf A and $\text{Path}^*(s, A)$ the set of finite paths starting in s

A trace is a finite or infinite sequence of external actions that is obtained from a path by omitting the states, internal actions and distributions. Let *trace*(B) denote the function that assigns to each execution its trace

Example

$$\left(\varepsilon\ \mathsf{snd}(1)\ \mu^{1}_{100}\ \varepsilon\ \mathsf{snd}(1)\ \mu^{99}_{100}\ 1\right)$$

is a finite path in the lossy channel example. It's trace is

$$(\mathsf{snd}(1)\ \mathsf{snd}(1))$$

Trace Distributions

- Semantics of a PA is defined by it's trace distributions
- Each trace distribution represents one of the potential visible behaviours
- Non-determinism is resolved by *randomized, partial and history dependent adversary*
- An adversary can be considered as the PA analog of an execution in an NA

Resolving Non-Determinism

- replace non-deterministic choices by *externally* defined probabilistic ones using an adversary
- e.g. in the channel example, how the channel is used.
- An adversary may be
  - partial (interrupt the execution at any time),
  - randomized (determine the choices randomly)
  - time-dependent (may base their choices causally)

Adversaries

Let PA A, $s_x \in S_A$ and adversary of A starting from s_x is a function

$$E : \text{Path}^*(s_x, A) \to D(A_A \times D(S_A) \cup \{\perp\})$$

where $\perp$ is the error or interruption state and such that if

$$E(\pi)(a, \mu) > 0 \Rightarrow \text{last}(\pi) \xrightarrow{a} \mu$$

Given a path π and a state $s \in S$ the adversary schedules transition $s \xrightarrow{a} \mu$ with probability

$$E(\pi)(a, \mu)$$

Example: Adversaries on the lossy channel PA

Let E_1:

$$E_1(\pi)(\mathsf{snd}(1), \mu^1_{100}) = 1 \qquad \text{if } \mathsf{last}(\pi) = \varepsilon$$

$$E_2(\pi)(\mathsf{rec}(i), \{\varepsilon \mapsto 1\}) = 1 \qquad \text{if } \mathsf{last}(\pi) = i$$

Let $i = 1, j \in \{100, 200\}$, E_2 :

$$E_1(\pi)(\mathsf{snd}(i), \mu^i_j) = \frac{1}{4} \qquad \text{if } \mathsf{last}(\pi) = \varepsilon$$

$$E_2(\pi)(\mathsf{rec}(i), \{\varepsilon \mapsto 1\}) = 1 \qquad \text{if } \mathsf{last}(\pi) = i$$

Adversaries and Markov Chains

Describe Adversary E starting in state s_0 as tree:

- root is s_0
- nodes are the finite paths, leaves are sequences $\pi\perp$ with $E(\pi)(\perp) > 0$.
- children of a node are suffixes of parent node $\pi a \mu t$, with $E(\pi)(a, \mu) > 0$ and $\mu(t) > 0$ and $\pi\perp$ if $E(\pi)(\perp) > 0$
- the edge from π to $\pi a \mu t$ is labeled with probability $E(\pi)(a, \mu) \cdot \mu(t)$

In fact this is a cycle-free discrete time Markov chain.

Paths, maximal Paths, Probability function for Paths in E

A path in an adversary E is a (probabilistic) path $\pi = s_0 a_1 \mu_1 \dots$ with

$$\forall 0 \leq i \leq |\pi| : E(s_0 a_1 \mu_1 \dots a_i \mu_i s_{i+1})(a_{i+1}, \mu_{i+1}) > 0$$

The maximal paths in E are the infinite paths π in E where $E(\pi)(\perp) > 0$. Denote $\text{Path}^{\text{max}}(E)$ as the set of all maximal Paths

Let π be a Path. Then the probability of π $Q^E(\pi)$ is defined by a recursive function:
Let PA A, $s \in S$. Then we define $Q^E(s) : \text{Path}^*(s, A) \to [0, 1]$ inductively by:

$$Q^E(s) = 1 \text{ and } Q^E(sa\mu t) = Q^E(s) \cdot E(s)(a, \mu) \cdot \mu(t)$$

Example

Reconsider E_1 and E_2. The path $\pi = \left(\varepsilon, \mathsf{snd}(1), \mu^{1}_{100}, \varepsilon, \mathsf{snd}(1), \mu^{99}_{100}, 1\right)$ is a path in E_1 and E_2. The following probabilities are assigned by the adversaries:

$$Q^{E_1}(\pi) = 1 \cdot \frac{1}{100} \cdot 1 \cdot \frac{99}{100} \qquad Q^{E_1}(\pi) = \frac{1}{4} \cdot \frac{1}{100} \cdot \frac{1}{4} \cdot \frac{99}{100}$$

Probability spaces

A probability space is a triple $(\Omega, \mathcal{F}, P)$ where

1. Ω is a set, called the sample space
2. $\mathcal{F} \subseteq 2^{\Omega}$ is a set of subsets of Ω which is closed under countable union and complement and which contains Ω. (σ-Algebra)
3. $P : \mathcal{F} \to [0, 1]$ is a probability measure on $\mathcal{F}$ which means $P(\Omega) = 1$ and for any countable pairwise disjoint subsets $\{X_i\}_i$ of $\mathcal{F}$: $P(\cup_i X_i) = \sum_i P(X_i)$

Probability space of adversaries

The probability space of a partial adversary E starting in $s \in S$ is defined by:

1. $\Omega_E = \text{Path}^{max}(E)$
2. $\mathcal{F}_E$ the smallest σ-algebra containing the set $\{C_\pi | \pi \in \text{Path}^*(E)\}$ with $C_\pi = \{\pi' \in \Omega_E | \pi \text{ prefix of } \pi'\}$
3. P_E the unique measure on $\mathcal{F}_E$ such that $\forall \pi \in \text{Path}^*(s, A) : P_E(C_\pi) = Q^E(\pi)$

Examples

Example 3.13 The collection $\mathcal{F}_E$ contains many sets of traces that occur in practice, or that come easily to one's mind. For instance, the set of paths containing at most three elements a is given by

$$\bigcup_{\rho \in X} C_\rho,$$

where $X = \{\alpha \in \mathit{Path}^*(\mathcal{A}) \mid \alpha \text{ contains at most three } a\text{'s}\}$. Since X is countable, the set above is an element of $\mathcal{F}_E$. The set containing the single infinite path π equals

$$\bigcap_{\rho \sqsubseteq \pi, \rho \neq \pi} C_\rho.$$

Examples

Example 3.14 Consider the adversary E_2 from Example 3.6. Then Ω_{E_2} is just the sets of all infinite paths. The set C_π contains the infinite paths extending the path π and $\mathcal{F}_{E_2}$ is the smallest σ–algebra containing those cones. Some values of the function $\mathbf{P}^{E_2}$ are

$$\mathbf{P}^{E_2}[C_{\langle \varepsilon\ snd(0)\ \mu^1_{100}\ 1\rangle}] = \mathbf{Q}^{E_2}(\langle \varepsilon\, snd(0)\mu^1_{100} 1\rangle) = \tfrac{1}{4} \cdot \tfrac{99}{100}.$$

and

$$\begin{aligned}
&\mathbf{P}^{E_2}[\text{a max. path generated by } E \text{ contains at most three actions } snd(0)] \leq \\
&\mathbf{P}^{E_2}[\text{a max. path generated by } E \text{ contains finitely many actions } snd(0)] = \\
&\mathbf{P}^{E_2}\left[\bigcup_i \text{a max. path generated by } E \text{ contains no } snd(0) \text{ after position } i\right] = \\
&\lim_{i\to\infty} \mathbf{P}^{E_2}[\text{a max. path generated by } E \text{ contains no } snd(0) \text{ after position } i] = \\
&\lim_{i\to\infty} 0 = 0.
\end{aligned}$$

The third step in this computation follows easily from the definition of probability space.

Trace distributions

Let $A_A^* = \{a \in A_A | a \in \text{Path}^*(E)\}$ and $A_A^\infty = \{a \in A_A | a \in \text{Path}^{\text{max}}(E)\}$

The trace distribution H of an adversary E, denoted by $trdistr(E)$ is the probability space defined by:

1. $\Omega_H = A_A^* \cup A_A^\infty$
2. $\mathcal{F}_H$ is the smallest σ-algebra that contains the sets $\{C_\beta | \beta \in A_A^*\}$ where $C_\beta = \{\beta' \in \Omega_H | \beta \text{ prefix of } \beta'\}$
3. P_H is defined by $\forall X \in \mathcal{F}_H : P_H(X) = P_E(\{\pi \in \Omega_E | \text{trace}(\pi) \in X\})$

We denote the set of all possible trace distributions of A by trdistr(A)

Examples

Example 3.16 Consider the trace distribution H of adversary E_2 from Example 3.6 again. The sets Ω_H and $\mathcal{F}_H$ need no further explanation. The probability on the set $\{\pi \mid trace(\pi) = snd(1)\}$, i.e. the maximal paths whose trace starts with $snd(1)$, is given as follows.

$$\begin{aligned}
\mathbf{P}_H[C_{snd(1)}] = \mathbf{P}_{E_2}&[C_{\langle \varepsilon snd(1) \mu^1_{100} 1\rangle}] \\
&+ \mathbf{P}_{E_2}[C_{\langle \varepsilon snd(1) \mu^1_{100} \varepsilon\rangle}] \\
&+ \mathbf{P}_{E_2}[C_{\langle \varepsilon snd(1) \mu^1_{200} 1\rangle}] \\
&+ \mathbf{P}_{E_2}[C_{\langle \varepsilon snd(1) \mu^1_{200} \varepsilon\rangle}] \\
&= \tfrac{1}{4} \cdot \tfrac{1}{100} + \tfrac{1}{4} \cdot \tfrac{99}{100} + \tfrac{1}{4} \cdot \tfrac{1}{200} + \tfrac{1}{4} \cdot \tfrac{199}{200} = \tfrac{1}{2}
\end{aligned}$$

Markov Chains

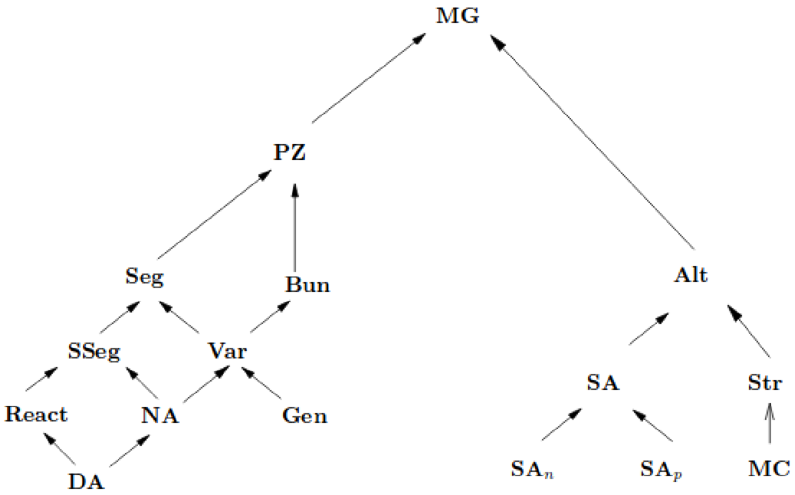
Discrete time Markov Chain (DTMC) is basically an unlabelled PA in which each state has exactly one one outgoing probabilistic transition.

Continuous time Markov Chain (CTMC) can be seen as DTMC in which each state s is assigned a rate $\lambda_s \in \mathbb{R}_{>\not{0}}$. The rate λ_s determines the sojourn time in s: The probability to stay in s for t time units is

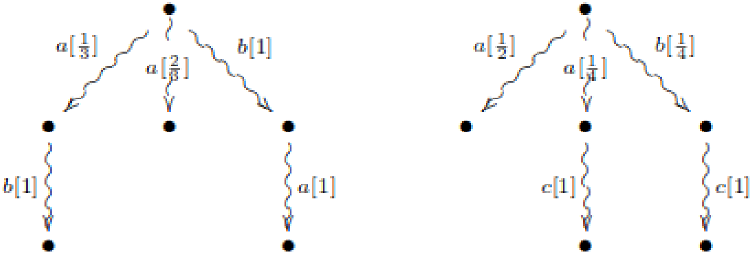
$$t = 1 - e^{\lambda_s \cdot t}$$

Markov decision Process (MDP) is a PA without internal actions in which each state contains at most one outgoing transition labelled with *a*

The PA Zoo [4]

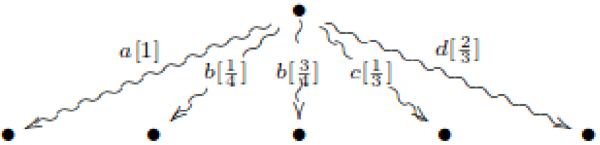


Example: Reactive, Generative [4]



Reactive system   Generative system

Example: I/O [4]



transitions from a state in an I/O probabilistic automaton
$A^{in} = \{a, b\}, \; A^{out} = \{c, d\}$

References I

D. Lehmann and M. O. Rabin, "On the advantages of free choice: A symmetric and fully distributed solution to the dining philosophers problem," in *Proceedings of the 8th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*, ACM, 1981, pp. 133–138.

K. P. Körding and D. M. Wolpert, "Bayesian integration in sensorimotor learning," *Nature*, vol. 427, no. 6971, p. 244, 2004.

T. S. Lee and D. Mumford, "Hierarchical bayesian inference in the visual cortex," *JOSA A*, vol. 20, no. 7, pp. 1434–1448, 2003.

A. Sokolova and E. P. De Vink, "Probabilistic automata: System types, parallel composition and comparison," in *Validation of Stochastic Systems*, Springer, 2004, pp. 1–43.

M. Stoelinga, "An introduction to probabilistic automata," ,