DISCRETE PROBABILISTIC SYSTEMS

Formal Models SS18

Martina Seidl
Institute for Formal Models and Verification

JOHANNES KEPLER UNIVERSITY LINZ

Motivation

challenge: capture random phenomena
$\Rightarrow$ models for probabilistic systems

examples

- randomized algorithms
  e.g., leader election, consensus
- modeling of unreliable/unpredictable system behavior
  e.g., message loss, garbling
- model-based performance evaluation
  e.g., distribution of message transmission delay, failure rate of a processor

Probabilities Revisited

terminology

- **sample space** Ω: possible outcomes of an experiment
- **event**: subset of Ω
- **probability** P: likelihood that an event occurs

examples

- toss a fair coin: $\Omega = \{h, t\}$
  events: $H = \{h\}$ (head), $T = \{t\}$ (tail), $P(H) = P(T) = 0.5$
- toss two fair coins: $\Omega = \{(h, h), (h, t), (t, h), (t, t)\}$
  event E: at least one h, $P(E) = 0.75$

Example: Coins and Dices

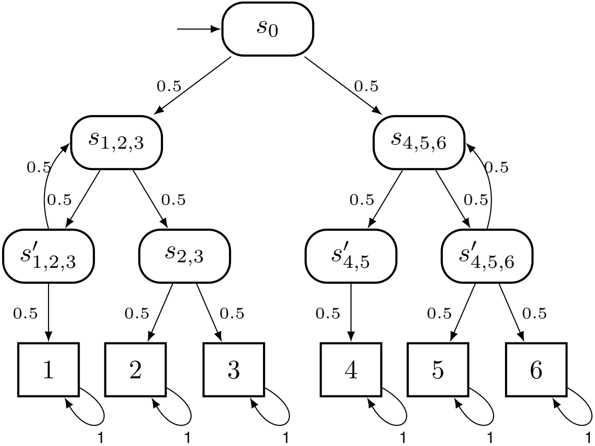
aim: model a six-sided die by tossing a fair coin

Example: Coins and Dices

aim: model a six-sided die by tossing a fair coin
(algorithm by Knuth/Yao)

algorithm:

- start in state s_0
- repeat until the value is decided
  - toss the coin
  - if h: left branch
  - if t: right branch



Markov Chain

Definition A Markov chain $M = (S, \iota, P)$ consists of the following components:

- a set of states S
- initial distribution $\iota \colon S \to [0,1]$ with $\sum_{s \in S} \iota(s) = 1$
- transition probability function $P \colon S \times S \to [0,1]$ with for all $s \in S$: $\sum_{s' \in S} P(s, s') = 1$

alternative representations: ι as vector and P as matrix

Properties of Markov Chains

- state-transition system (no labels!) augmented with probabilities
- no memory: In the current state, the future states are independent of the past states
- every state has at least one outgoing transition
- states s with $\iota(s) > 0$ are possible initial states
- states s' with $P(s, s') > 0$ are possible successors of s
- absorbing state s: $P(s, s) = 1$ and for all $s \neq s'$: $P(s, s') = 0$
- for $T \subseteq S$: $P(s, T) = \sum_{t \in T} P(s, t)$

LTS vs Markov Chain

- LTS
  - (S, I, Σ, T) with $T \subseteq S \times \Sigma \times S$
  - non-deterministic behavior

- Markov chain
  - (S, ι, P) with $P\colon S \times S \to [0, 1]$
  - probabilistic behavior

How to combine probabilistic and non-deterministic behavior?

$\Rightarrow$ **Markov Decision Processes (MDP)**

Motivation MDP

- probabilistic and non-deterministic choices in one model
  - randomized distributed algorithms
  - quantify outcomes of randomized actions
- formalization of agent-environment interaction
  - agent's utility depends on sequence of decisions
  - produce optimal behavior that balances risks and rewards of acting in uncertain environment

Markov Decision Process

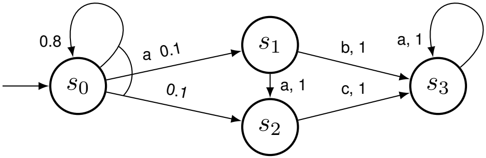
Definition A Markov Decision Process $M = (S, \iota, \Sigma, T, R, \gamma)$ consists of the following components:

- set of states S
- initial distribution $\iota\colon S \to [0,1]$ with $\sum_{s \in S} \iota(s) = 1$
- transition probability function $T\colon S \times \Sigma \times S \to [0,1]$ with for all $s \in S$, $a \in \Sigma$: $\sum_{s' \in S} T(s, a, s') \in \{0, 1\}$
- reward function $R\colon S \times \Sigma \to \mathbb{R}$
- discount factor γ

Properties of MDP

- action a is enabled in state s iff $\sum_{s' \in S} T(s, a, s') = 1$
- every state has at least one enabled action
- MDP is MC iff each state of MDP has exactly one enabled action
- intuitive operational behavior:
  - a starting state s_0 is entered according to ι by a stochastic experiment
  - when a state s is entered
    1. an enabled action $a \in \Sigma$ is selected
    2. a successor state s' with $T(s, a, s') \neq 0$ is selected randomly according to distribution $T(s, a, *)$
    3. s' is entered
- for $S' \subseteq S$, $a \in \Sigma$: $T(s, a, S') = \sum_{s' \in S'} T(s, a, s')$

Example (without rewards)



- $S = \{s_0, s_1, s_2, s_3\}$
- $\iota(s_0) = 1$
- $\Sigma = \{a, b, c\}$
- $T(s_0, a, s_0) = 0.8, T(s_0, a, s_1) = 0.1, T(s_0, a, s_2) = 0.1$
  $T(s_1, b, s_3) = 1, T(s_1, a, s_2) = 1$
  $T(s_2, c, s_3) = 1$
  $T(s_3, a, s_3) = 1$

Rewards

idea: maximize the cumulative reward in the long run

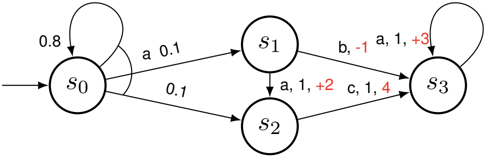
Definition Let $R_{t+1}, R_{t+2}, \ldots$ be the sequence of rewards received after time step t. The expected return at t, denoted by G_t, is a specific function of the reward sequence like

$$G_t = R_{t+1} + R_{t+2} + \ldots + R_T$$

where

- T is the final time step for episodic tasks (e.g., a game)
- $T = \infty$ for continuous tasks with infinite horizons (e.g., robot with long life span)

Example (with rewards)



- $S = \{s_0, s_1, s_2, s_3\}$
- $\iota(s_0) = 1$
- $\Sigma = \{a, b, c\}$
- $T(s_0, a, s_0) = 0.8, T(s_0, a, s_1) = 0.1, T(s_0, a, s_2) = 0.1$
  $T(s_1, b, s_3) = 1, T(s_1, a, s_2) = 1$
  $T(s_2, c, s_3) = 1$
  $T(s_3, a, s_3) = 1$
- $R(s_3, a) = 3, R(s_1, a) = 1, R(s_1, b) = -1, R(s_2, c) = 4$

Discounts

idea: realize a more farsighted agent $\Rightarrow$ discounting

a reward might be more valuable now than in the future, because of

- inflation
- obliteration

Definition The discounted return is defined by

$$G_t = R_{t+1} + \gamma R_{t+2} + \ldots + \gamma^2 R_{t+3} + \ldots$$

Policies

Given an agent who wants to achieve a goal.

Question: what is the solution to achieve the goal?

- a fixed sequence of actions is not a solution
- a solution is a policy π that specifies what the agent should do in any state that it reaches
- if a policy is complete, the agent will always know what to do next
- a policy is optimal, if it yields the highest expected utility

Literature

Christel Baier, Joost-Pieter Katoen:
Principles of Model Checking
MIT Press, 2008

Richard S. Sutton, Andrew G. Barto:
Reinforcement Learning – An Introduction
Adaptive computation and machine learning, MIT Press, 1998

Stuart J. Russell, Peter Norvig:
Artificial Intelligence – A Modern Approach
Pearson Education, 2010