Markov decision processes

Quantitative Logics

David N. Jansen

Repetition...

- Kripke structure / transition system:
  choice between transitions is
  nondeterministic
- Markov chain:
  choice between transitions is
  (fully) probabilistic

- Can I have both?

Let’s play a game

Let’s play a better game

What do we need?

- simple game:
  move = probabilistic choice
- interesting game:
  move = tactical choice + probabilistic choice
  - tactical choice done by external entity
  - based on current situation
  - we cannot or want not
    assign probabilities to tactical choices
  - assumptions: tactical choice first. one player only

Markov decision process

- In a MDP, transitions combine nondetermistic and probabilistic choices!

- Examples:
  - one-player strategy games (Patience)
  - algorithms where some parts are random

Formal definition

- A Markov decision process consists of:
  - S — finite set of states (often $S = \{1, 2, ... n\}$)
  - Act — finite set of actions
  - $\mathbf{P}: S \times Act \times S \rightarrow [0,1]$ — transition probability matrix for every action
  - $\pi_0: S \rightarrow [0,1]$ — initial state distribution (sometimes)

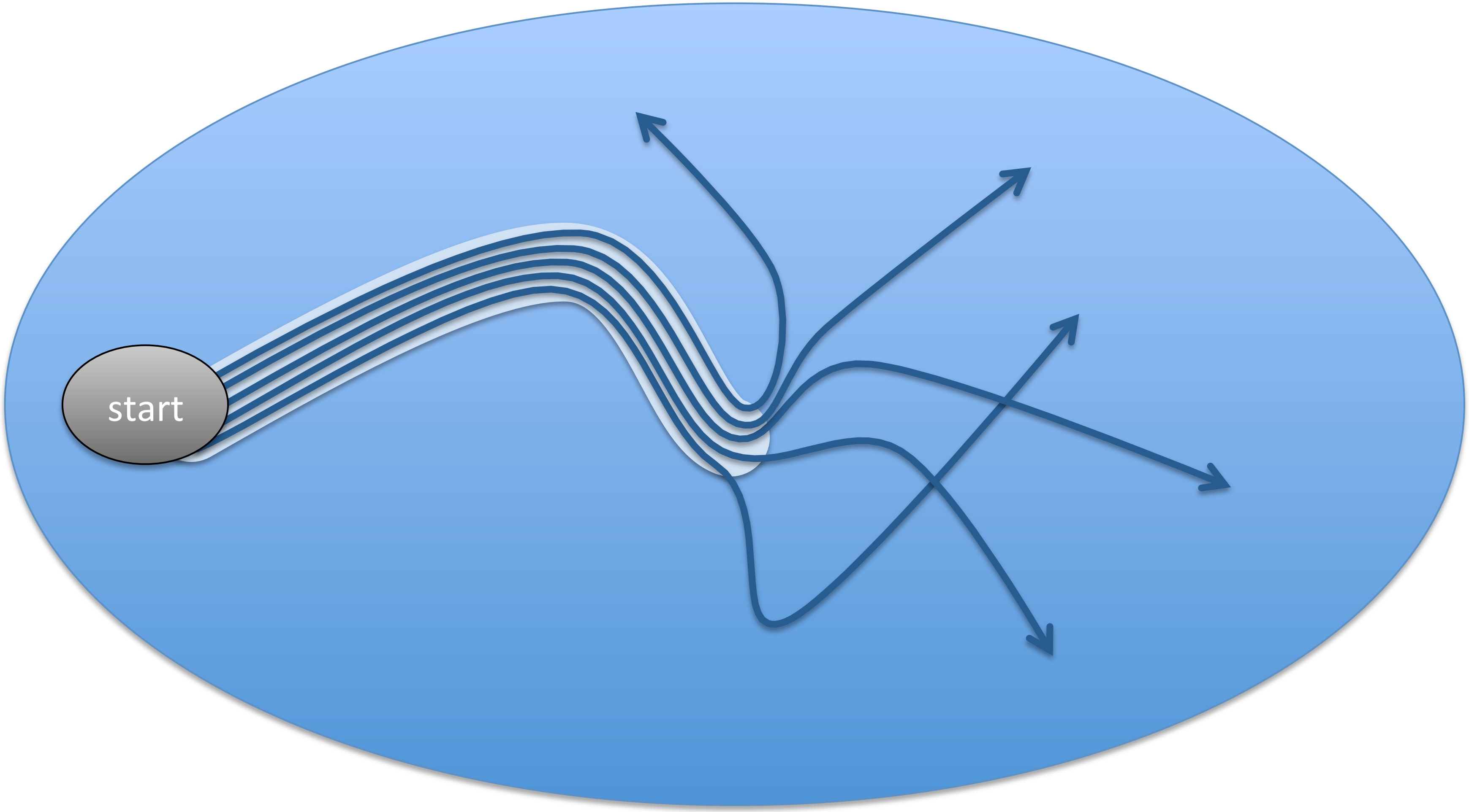
Semantics of a Markov Decision Process

- similar to Kripke structure:
  - system starts in one of the initial states, chosen according to π_0
  - system is always in a state
  - from time to time, a transition is taken:
    - some external entity chooses an action α
    - when the system leaves state i, the next state is j with probability $\mathbf{P}(i,\alpha,j)$

Example: Yahtzee

- first roll: player rolls all five dice
- later: player chooses 0–5 dice to roll again
- some combinations of dice give points
  - Pair, Triple, Carré, Yahtzee: 2–5 equal faces
  - Full House: Triple + Pair
  - 1, 2, ..., 6: any die with that face counts
  - etc.

Recall:
Cylinder Set of a Markov Chain



Probability Space of a Markov Chain

- Ω all paths
- $\mathcal{F}$ σ-algebra generated by cylinder sets
  - $\text{Cyl}(s_0, s_1, ..., s_n)$:= paths starting with $s_0, s_1, ..., s_n$
  - complements and unions of cylinder sets
- μ unique extension of
  $\mu(\text{Cyl}(s_0, s_1, ..., s_n))$
  $= \pi_0(s_0) \cdot \mathbf{P}(s_0,s_1) \cdot \mathbf{P}(s_1,s_2) \cdots \mathbf{P}(s_{n-1},s_n)$

No unique probability space!

- Probability to take a transition depends on external choices

- To resolve nondeterminism, introduce a scheduler/policy/strategy/adversary.

Strategy

- A strategy is a mapping from paths in the MDP to actions.

- may depend on full history!
- in practice, often only depends on
  - last state
  - last state + length of history
  - last state + finite „memory"

Example: I already had a full house, so I should try something else.

Induced Markov chain

- A MDP + a strategy together induce a Markov chain.
- MC may be infinite.

- There is a one–one correspondence between finite paths of the MDP + strategy and paths of the induced MC.

Probability space of a MDP + a strategy

- The probability space of the induced MC is a probability space of the MDP.
- Notation: $Prob^{\sigma}(Cyl(s_0,s_1,\ldots))$ for strategy σ

interesting measures

- Find the least/highest probability to reach a state

PCTL (interpreted on MDPs)

- Find the best/worst strategy

What to do with unused actions?

possible ways of handling:

- self loop
- error state
- join with a real choice

- (allow that not every action can be chosen, e.g. by a function that tells which actions are allowed in a state)

Randomised mutual exclusion

- Two processes compete for a resource.
- If both want access at the same time, an arbiter picks one process at random.
- Otherwise, every process gets access whenever it tries.

- Draw a MDP for this situation!

Recapitulation

- Markov decision processes combine nondeterministic and probabilistic choices.
- Strategies select how to resolve nondeterministic choices.
- MDP + strategy induce a Markov chain.
- Probabilities on a MDP are defined via the induced Markov chain.