Aalto University
School of Science

Markov chains and Markov decision processes

Emil af Björkesten

Presentation 5

2.10.2020

MS-E2191 Graduate Seminar on Operations Research
Fall 2020

Snakes and ladders

Will I ever win?

How long will it take?

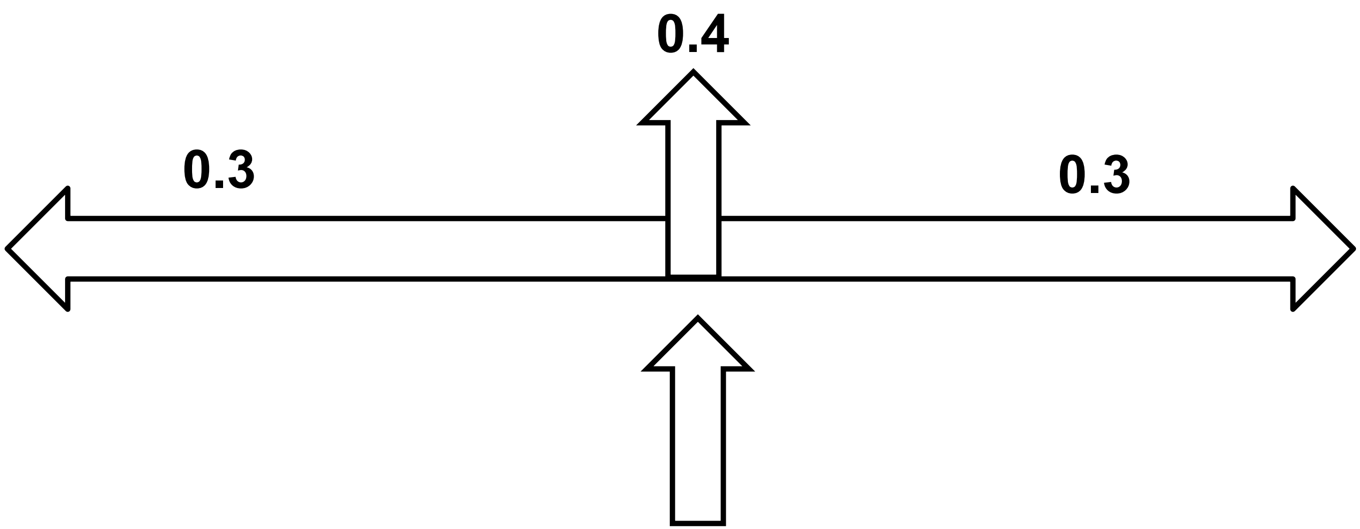
Where will I be after 10 turns?

https://fun-play.co.uk/shop/fun-games/snakes-ladders-1-100-solid/

Stock prices

GOOG stock price development, from finance.yahoo.com

Where do we end up?



Stochastic processes

A process where transitions between states are stochastic

Inventory management

Weather models

Natural languages

Queuing

Markov chain

Memoryless process

Transition probabilities and initial state known

The current state is dependent on the previous one:

$$P(X_{t+1}) \neq P(X_{t+1} \mid X_t)$$

Transition probabilities

today / tomorrow	Sunny	Cloudy	Rainy
Sunny	0	0.5	0.5
Cloudy	0.2	0.6	0.2
Rainy	0.3	0.4	0.3

$$P(w_{t+1} = Sunny \mid w_t = Sunny) = 0$$

States and transition probabilities

Row sum = 1

$$\boldsymbol{P} = \begin{bmatrix} \mathbf{0} & \mathbf{0.5} & \mathbf{0.5} \\ \mathbf{0.2} & \mathbf{0.6} & \mathbf{0.2} \\ \mathbf{0.3} & \mathbf{0.4} & \mathbf{0.3} \end{bmatrix}$$

Distribution at next time step

$$\mu_{t+1} = \mu_t P$$

$$\mu_{t+n} = \mu_t P^n$$

Terminology

Frequency

$$N_t(y) = \sum_{s=0}^{t} \mathbf{1}(X_s = y)$$

Occupancy

$$M_t(x, y) = \mathbb{E}\left(N_y(t) \middle| X_0 = x\right)$$

$$M_t = \sum_{s=0}^{t} P^s$$

Periodicity

Irreducibility: no isolated states

Snakes and ladders

Will I ever win?

How long will it take?

Where will I be after 10 turns?

https://fun-play.co.uk/shop/fun-games/snakes-ladders-1-100-solid/

Invariant distribution

End state (if the system has one): $\lim_{t\to\infty} \mu_t$

Limiting distribution = invariant distribution $\pi P = \pi$

May depend on the initial state

Simulation

Is the process really memoryless?

Is a discrete-time model motivated?

- Continuous-time process: add a random time component

Find out the probabilities for each transition

Markov chains for language processing

Empirical transition probabilities can be used to produce strings of words (fiction, chatbots, gene sequences)

https://github.com/StrikingLoo/ASOIAF-Markov uses the A Song of Ice and Fire books to produce new sentences (and write the sixth book?)

```
stochastic_chain('The bold')
'The bold ones have had those formalities of greeting . " Asha asked her how it was'
```

```
stochastic_chain('Jon Snow')
'Jon Snow smiled . " There was something foul; the heaving grey - green eyes . She'
```

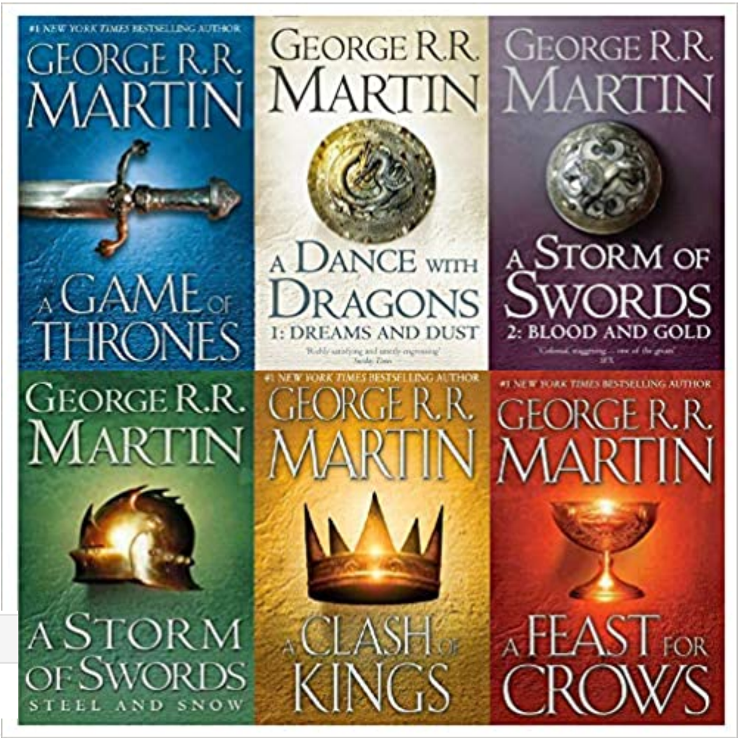
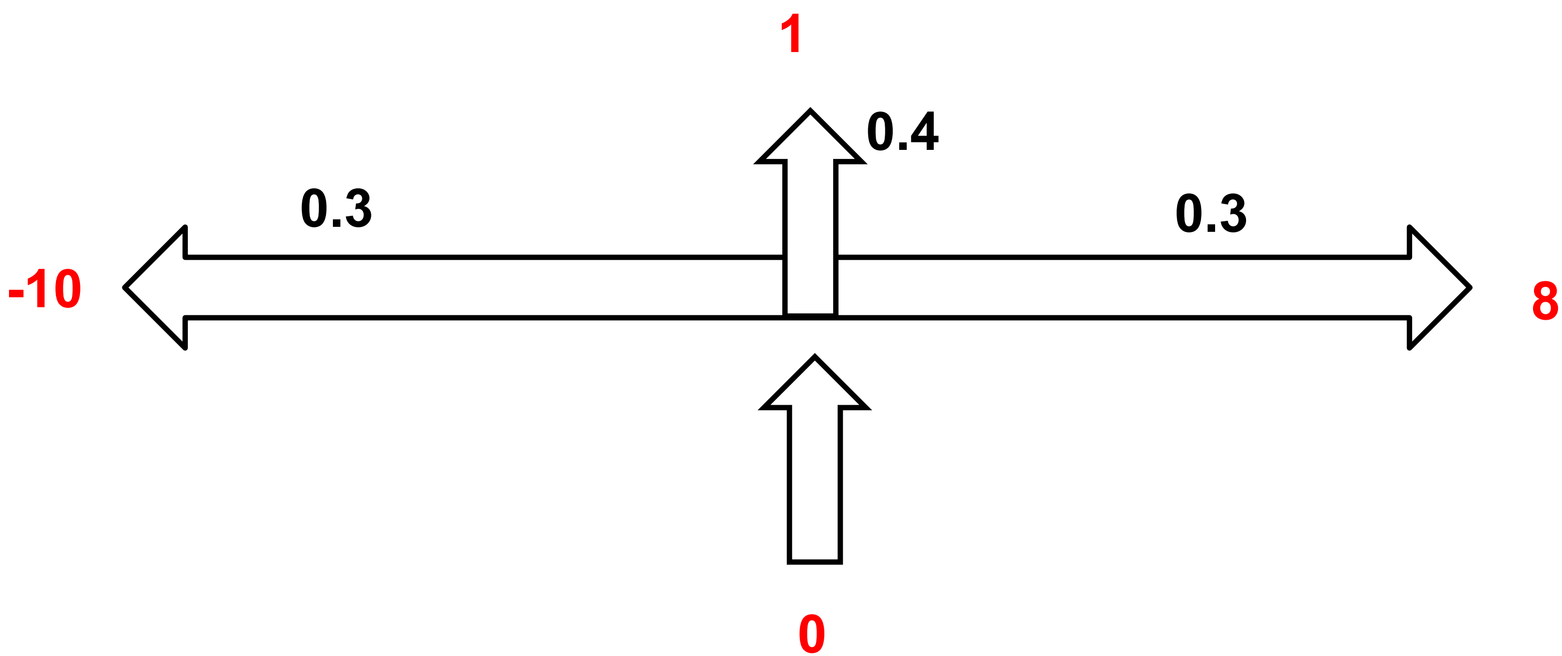


Image: https://www.amazon.com/Thrones-Collection-George-Martin-Dragons/dp/9369763740

Should we stay or go?



Markov decision processes

Some states are better than others

Actions lead to states with some probability distribution

Wanting to find the best possible policy

Associate each transition with a reward

Using MDPs

Define reward functions

Solve using dynamic algorithm

Find the best possible decision rules and policies

$$v(s) = \sum_{t \in S} P(s,a,t)(R(s,a,t) + fv(t))$$

v **value;** S **states;** s **current state;** a **chosen action;** f **discount factor;** R **reward;** P **probability**

Maximizing rewards

Which costs/rewards are significant?

What to maximize?

- Discounted sum
- Average reward
- Total reward

Policy choice: What to base the policy on?

Is the state completely observed?

Solving it

Find the optimal action for each state

$$v(s) = \max_{a \in A} \sum_{t \in S} P(s,a,t)\big(R(s,a,t) + f v(t)\big)$$

Applications

Inventory management

Road maintenance

Image: https://www.kimble.fi/

Summary

A *Markov chain* consists of *states* and *transition probabilities*

Memoryless, same transition probabilities for each time step

A *Markov decision process* has states, *actions*, transition probabilities, and *rewards*

Optimize the reward using some criterion

References

Leskelä, L. (2018). Stokastiset prosessit (lecture material). https://math.aalto.fi/~lleskela/papers/Leskela_2018-08-07_Stokastiset_prosessit.pdf

Puterman, M. L. (1994). Markov Decision Process: Discrete Stochastic Dynamic Programming. John Wiley & Sons

Silver, D. (2020). Markov Decision Processes (lecture material). https://www.davidsilver.uk/wp-content/uploads/2020/03/MDP.pdf

Strika, L. (2019). ASOAIF-Markov (GitHub repository.) https://github.com/StrikingLoo/ASOIAF-Markov

Homework: Motion on a grid

Each action taken will have the desired outcome with P=0.7 (if the desired outcome is possible). Transitions to all other neighbouring cells and not moving at all are equally likely outcomes. Diagonal motion is not possible.

At location 2, choosing action EAST

P(3) = 0.7

P(2) = 0.1

P(1) = 0.1

P(5) = 0.1

7	8	9
4	5	6
1	2	3

https://maps.google.com

Homework: Motion on a grid

Each action taken will have the desired outcome with P=0.7 (if the desired outcome is possible). Transitions to all other neighbouring cells and not moving at all are equally likely outcomes. Diagonal motion is not possible.

At location 2, choosing action SOUTH

P(1) = 0.25

P(2) = 0.25

P(3) = 0.25

P(5) = 0.25

7	8	9
4	5	6
1	2	3

https://maps.google.com

Homework: Motion on a grid

We choose NORTH as our policy.

1. **Starting at 1, what is the distribution at t=10?**
2. **Is there an invariant distribution, and does it depend on the initial state?**
3. **The reward for a transition to 9 is 100, all other transitions have reward 0. Use the equation on slide 17 with discount factor 0.5. What is the value of each state?**

7	8	9
4	5	6
1	2	3

Send to emil.afbjorkesten@aalto.fi, DL 9.10