Death and Discounting

COMP 767

Pascale Gourdeau

Material from: Adam Schwartz. *Death and Discounting.* Technion – Israel Institute of Technology. (2000)

April 7th, 2017

Overview

Motivation

Question: Can MDPs that exhibit phenomena on several different time scales be modelled through an objectivee function which is a linear combination of several discounted cost functions?

Main Results

1. Discounted MDPs with total-cost MDP with finite random duration are equivalent (restated).

Main Results

1. Discounted MDPs with total-cost MDP with finite random duration are equivalent (restated).
2. MDPs with several time scales and MDPs with several discount factors are not equivalent.

Main Results

1. Discounted MDPs with total-cost MDP with finite random duration are equivalent (restated).
2. MDPs with several time scales and MDPs with several discount factors are not equivalent.
3. Mixed discount MDPs are equivalent to MDPs with several time scales when the end of one time scale cannot be detected.

An Undiscounted finite-time MDP

Given an MDP with discount factor β.

1. Add an absorbing state Δ with zero cost.

An Undiscounted finite-time MDP

Given an MDP with discount factor β.

1. Add an absorbing state Δ with zero cost.
2. At each state, with probability $1 - \beta$ there is a transition to Δ.

An Undiscounted finite-time MDP

Given an MDP with discount factor β.

1. Add an absorbing state Δ with zero cost.
2. At each state, with probability $1-\beta$ there is a transition to Δ.
3. Let τ be the first time we reach the absorbing state

Equivalence

- This new undiscounted MDP can be seen as the original discounted MDP where we flip a coin after each action, where we either stay in the MDP or we move to the absorbing state.

Equivalence

- This new undiscounted MDP can be seen as the original discounted MDP where we flip a coin after each action, where we either stay in the MDP or we move to the absorbing state.
- This allows to have a policy independent of the coin flip (all actions are available in the absorbing state), and the same policy is well-defined in both models.

Equivalence

- This new undiscounted MDP can be seen as the original discounted MDP where we flip a coin after each action, where we either stay in the MDP or we move to the absorbing state.
- This allows to have a policy independent of the coin flip (all actions are available in the absorbing state), and the same policy is well-defined in both models.
- This undiscounted MDP is equivalent to the original, discounted-cost MDP.

Equivalence

- This new undiscounted MDP can be seen as the original discounted MDP where we flip a coin after each action, where we either stay in the MDP or we move to the absorbing state.
- This allows to have a policy independent of the coin flip (all actions are available in the absorbing state), and the same policy is well-defined in both models.
- This undiscounted MDP is equivalent to the original, discounted-cost MDP.
  - They have the same *expected return* for the same policy, and the same value function (except for the absorbing state).

Equivalence

- This new undiscounted MDP can be seen as the original discounted MDP where we flip a coin after each action, where we either stay in the MDP or we move to the absorbing state.
- This allows to have a policy independent of the coin flip (all actions are available in the absorbing state), and the same policy is well-defined in both models.
- This undiscounted MDP is equivalent to the original, discounted-cost MDP.
  - They have the same *expected return* for the same policy, and the same value function (except for the absorbing state).
- **Consequence:** Both models can be used for optimization.

Mixed Discount MDPs

We define two new MDPs:

- For M_1, let $I \in \mathbb{N}$ and for each $i = 1, \ldots, I$, we introduce a cost function c_i and its associated discount factor β_i.

Mixed Discount MDPs

We define two new MDPs:

- For M_1, let $I \in \mathbb{N}$ and for each $i = 1, \ldots, I$, we introduce a cost function c_i and its associated discount factor β_i.
- So we have I different value functions $V^i(x, \pi)$ given a certain policy.

Mixed Discount MDPs

We define two new MDPs:

- For M_1, let $I \in \mathbb{N}$ and for each $i = 1, \ldots, I$, we introduce a cost function c_i and its associated discount factor β_i.
- So we have I different value functions $V^i(x, \pi)$ given a certain policy.
- We get a total cost V as the sum $\sum_{i=1}^{I} V^i(x, \pi)$.

Mixed Discount MDPs

We define two new MDPs:

- For M_1, let $I \in \mathbb{N}$ and for each $i = 1, \ldots, I$, we introduce a cost function c_i and its associated discount factor β_i.
- So we have I different value functions $V^i(x, \pi)$ given a certain policy.
- We get a total cost V as the sum $\sum_{i=1}^{I} V^i(x, \pi)$.
- For M_2, for each cost function, we again have a coin flip as in the previous example, and associated random stopping time τ^i (to replace the discount factor) to.

(No) Equivalence

- There is no equivalence between M_1 and M_2.

(No) Equivalence

- There is no equivalence between M_1 and M_2.
- We can derive a counter example where $I = 2$ and the discount factors (coin flip probabilities) are not the same in M_1 (M_2).

(No) Equivalence

- There is no equivalence between M_1 and M_2.
- We can derive a counter example where $I = 2$ and the discount factors (coin flip probabilities) are not the same in M_1 (M_2).
- Then, it can be shown that the total value function under the same policy for M_2 is always greater than the one for M_1.

(No) Equivalence

- There is no equivalence between M_1 and M_2.
- We can derive a counter example where $I = 2$ and the discount factors (coin flip probabilities) are not the same in M_1 (M_2).
- Then, it can be shown that the total value function under the same policy for M_2 is always greater than the one for M_1.
- However, if the stopping times τ^i are not observable, the models are equivalent, provided the policies are not dependent on the coin flips (proved the same way as one discount factor).