Time Consistent Discounting

Tor Lattimore[1] Marcus Hutter[1,2]

October 7, 2011

[1]Australian National University

[2]ETH Zürich

Overview

Why study general discounting?

What is time-inconsistency and who cares?

What discount functions are time-(in)consistent?

How to act when using a time-inconsistent discount function?

Markov Decision Process

A Markov Decision Process (MDP) is a tuple $(\mathcal{S}, \mathcal{A}, T, R)$ where

1. $\mathcal{S}$ is a (possibly infinite) set of states.
2. $\mathcal{A}$ is the set of actions available to the agent.
3. $T(s'|s, a)$ is the probability that the agent travels from state s to s' given action a.
4. $R(s, a)$ is the reward given to the agent when it reaches state s having taken action a.

General Discounting

Define

1. A *policy* is a function $\pi : \mathcal{S} \to \mathcal{A}$
2. $R_\pi(s)$ is the expected sequence of rewards given to the agent when following policy π in state s.

How to choose the best policy?

Option 1. Maximise $\sum_{t=0}^{\infty} R_\pi(s)_t$. Might be infinite.

Option 2. Maximise $\sum_{t=0}^{\infty} \gamma^t R_\pi(s)_t$ where $\gamma \in (0, 1)$. Restrictive.

Option 3. Maximise $V^\pi_{\mathbf{d}^k}(s) := R_\pi(s) \cdot \mathbf{d}^k$ where k is current time-step and

$$\mathbf{d}^k \in [0, 1]^\infty \qquad \sum_{t=1}^{\infty} d_t^k = 1$$

Most general linear model.

Why not geometric?

1. May want growing farsightedness (unknown required horizon)
2. May know life-time of the agent
3. Learning is affected by the discount rate
4. Humans don't discount geometrically

Time Inconsistency

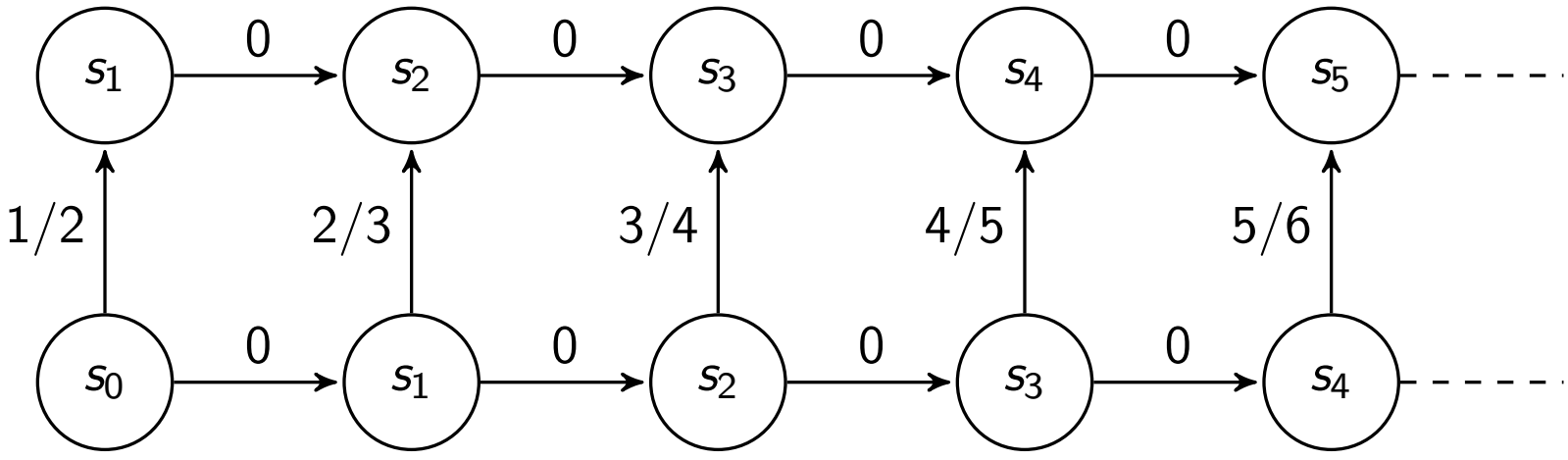
An agent in a known environment is time-inconsistent if it changes its plans for a future state over time.

Suppose that in one year an agent must choose between having \$100 or waiting an extra day for \$105. Its plausible that:

1. Initially it plans to wait for \$105
2. In one year it changes its mind and takes the immediate \$100

There are discount rates that make these actions "rational". We aim to classify such discount functions.

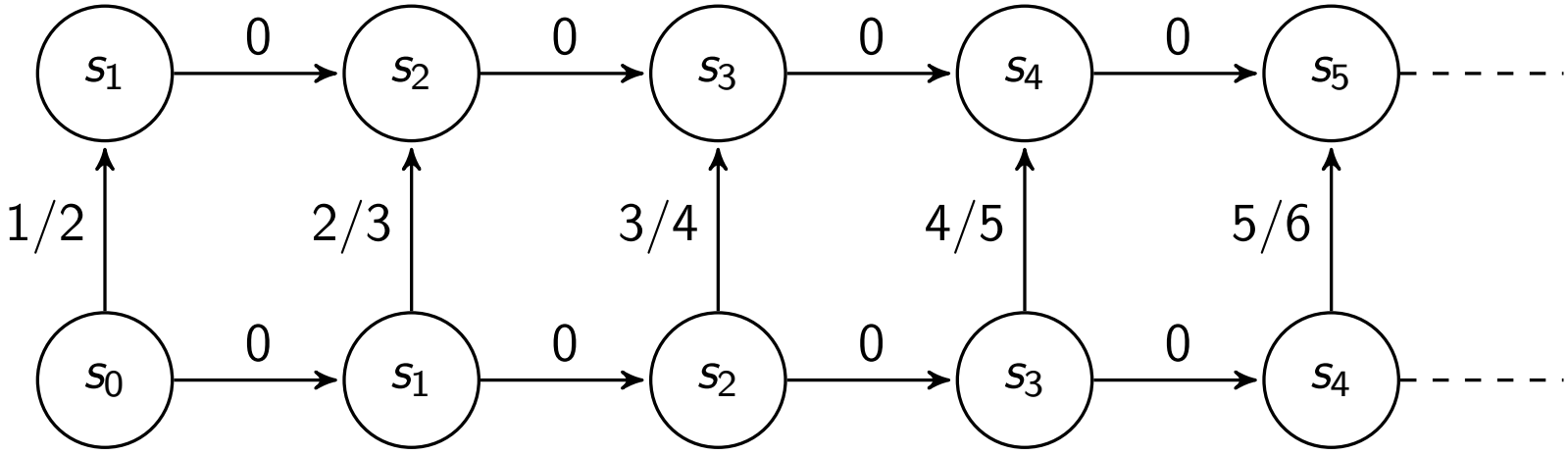
Example - Fixed Depth Failure



$$
\begin{aligned}
\mathbf{d}^0 &= \tfrac{1}{2} \quad \tfrac{1}{2} \quad 0 \quad 0 \quad 0 \quad 0 \quad \cdots \\
\mathbf{d}^1 &= \quad \tfrac{1}{2} \quad \tfrac{1}{2} \quad 0 \quad 0 \quad 0 \quad \cdots \\
\mathbf{d}^2 &= \quad \quad \tfrac{1}{2} \quad \tfrac{1}{2} \quad 0 \quad 0 \quad \cdots \\
\mathbf{d}^3 &= \quad \quad \quad \tfrac{1}{2} \quad \tfrac{1}{2} \quad 0 \quad \cdots
\end{aligned}
$$

Agent waits forever and is never rewarded.

Example - Geometric



$$
\begin{aligned}
\mathbf{d}^0 &= \gamma^0 \quad \gamma^1 \quad \gamma^2 \quad \gamma^3 \quad \gamma^4 \quad \gamma^5 \quad \dots \\
\mathbf{d}^1 &= \quad \gamma^1 \quad \gamma^2 \quad \gamma^3 \quad \gamma^4 \quad \gamma^5 \quad \dots \\
\mathbf{d}^2 &= \quad \quad \gamma^2 \quad \gamma^3 \quad \gamma^4 \quad \gamma^5 \quad \dots \\
\mathbf{d}^3 &= \quad \quad \quad \gamma^3 \quad \gamma^4 \quad \gamma^5 \quad \dots
\end{aligned}
$$

$$t_{up} = \arg\max_t \gamma^t (t+1)/(t+2)$$

New Results

Theorem (Characterization)

The following are equivalent

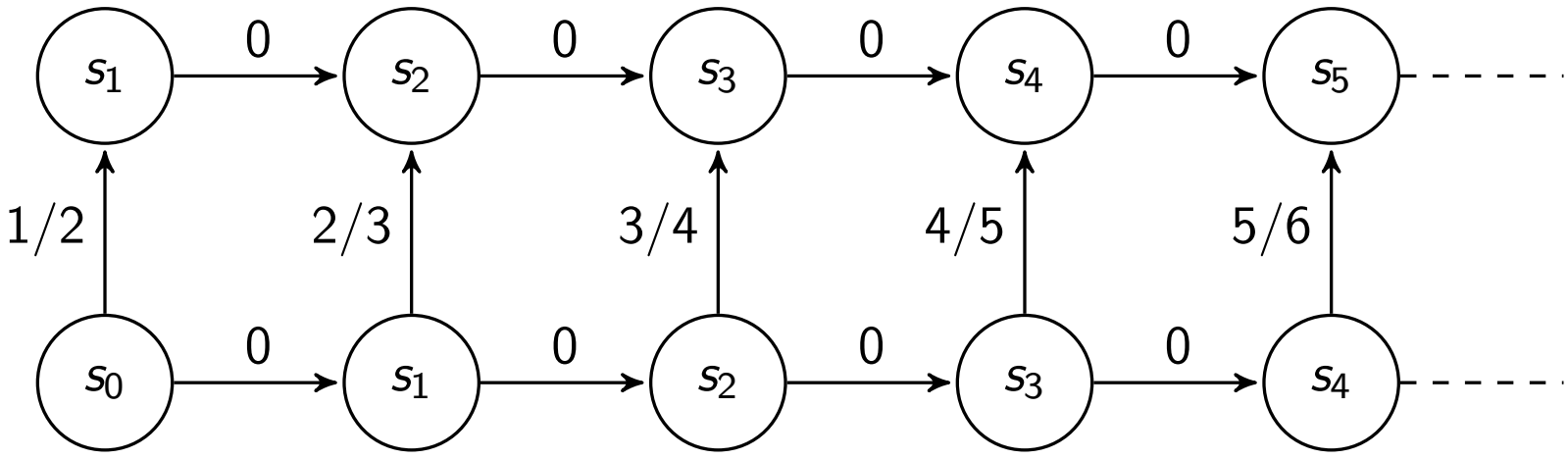
1. $\left\{\mathbf{d}^k\right\}$ *are time-consistent*
2. *For each* k *there exists a constant* $\alpha_k > 0$ *such that* $d_t^k = \alpha_k d_t^0$ *for all* $t \geq k$

Corollary (Strotz 1957)

If the same discount sequence is used at each time-step $(d_t^k = f(t-k))$ *then only geometric discounting is time-consistent.*

Example

There are non-geometric time-consistent discount rates



$$
\begin{array}{lcccccccc}
\mathbf{d}^0 & = & 1/1 & 1/4 & 1/9 & \cdots & 1/t^2 & 1/(t+1)^2 & \cdots \\
\mathbf{d}^1 & = & & 1/4 & 1/9 & \cdots & 1/t^2 & 1/(t+1)^2 & \cdots \\
\mathbf{d}^2 & = & & & 1/9 & \cdots & 1/t^2 & 1/(t+1)^2 & \cdots
\end{array}
$$

Example

Time Consistent

- Geometric: $d_t^k = \gamma^{k-t}$
- Fixed Life: $d_t^k = [\![t < H]\!]$
- Power: $d_t^k = 1/t^2$

Time Inconsistent

- Fixed Horizon: $d_t^k = [\![t - k < H]\!]$
- Hyperbolic, $d_t^k = 1/[1 + \kappa(t - k)]$ (popular in economics)

New Results

Theorem (Continuity)

Small perturbations in a time-consistent discount function cannot make it "seriously" time-inconsistent.

- Actual theorem statement is more technical.
- Essentially means we needn't worry about small rounding errors, even if they occur in all terms.

Game Theory

What to do if you know you're time inconsistent?

Treat your future selves as "opponents" in an extensive game.

Definition (Sub-game Perfect Equilibirum Policy)

A *sub-game perfect equilibrium policy* is a policy all players could agree on such that no player would wish to deviate from the plan.

Theorem

For any discount function $\mathbf{d}$ *there exists at least one sub-game perfect equilibrium policy.*

Similar to the idea of pre-commitment.

Problem! Sub-game perfect equilibrium policies are *not* unique.

Summary

- Introduced general discounting
- Classified time-consistent discount functions
- Showed values depend only slightly on perturbations in discount functions
- Showed the existence of "optimal" strategies for time-inconsistent discount functions