Continuous Markov Decision Processes with a Probability Theory Introduction

COMP 767

Pascale Gourdeau

March 17th, 2017

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite state of actions,

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite state of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite state of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,
- $P : S \times A \times \Sigma \to [0, 1]$ is a labelled stochastic transition kernel:

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite state of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,
- $P : S \times A \times \Sigma \to [0, 1]$ is a labelled stochastic transition kernel:
  - $\forall a \in A$, $\forall s \in S$, $P(s, a, \cdot) : \Sigma \to [0, 1]$ is a probability measure,

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite state of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,
- $P : S \times A \times \Sigma \to [0, 1]$ is a labelled stochastic transition kernel:
  - $\forall a \in A$, $\forall s \in S$, $P(s, a, \cdot) : \Sigma \to [0, 1]$ is a probability measure,
  - $\forall a \in A$, $\forall X \in \Sigma$, $P(\cdot, a, X) : S \to [0, 1]$ is a measurable function.

Measurable Space

Let Σ be a collection of subsets of S. We say that (S, Σ) is a *measurable space* if Σ is a σ-algebra, namely:

(i) $S \in \Sigma$,

Measurable Space

Let Σ be a collection of subsets of S. We say that (S, Σ) is a *measurable space* if Σ is a σ-algebra, namely:

(i) $S \in \Sigma$,

(ii) $A \in \Sigma \implies A^C \in \Sigma$,

Measurable Space

Let Σ be a collection of subsets of S. We say that (S, Σ) is a *measurable space* if Σ is a σ-algebra, namely:

(i) $S \in \Sigma$,

(ii) $A \in \Sigma \implies A^C \in \Sigma$,

(iii) If $A_n \in \Sigma$ for all $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

Measurable Space

Let Σ be a collection of subsets of S. We say that (S, Σ) is a *measurable space* if Σ is a σ-algebra, namely:

(i) $S \in \Sigma$,

(ii) $A \in \Sigma \implies A^C \in \Sigma$,

(iii) If $A_n \in \Sigma$ for all $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

Measurable Space

Let Σ be a collection of subsets of S. We say that (S, Σ) is a *measurable space* if Σ is a σ-algebra, namely:

(i) $S \in \Sigma$,

(ii) $A \in \Sigma \implies A^C \in \Sigma$,

(iii) If $A_n \in \Sigma$ for all $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

In probability theory,

- $S = \Omega$, the sample space – the set of all possible outcomes,

Measurable Space

Let Σ be a collection of subsets of S. We say that (S, Σ) is a *measurable space* if Σ is a σ-algebra, namely:

(i) $S \in \Sigma$,

(ii) $A \in \Sigma \implies A^C \in \Sigma$,

(iii) If $A_n \in \Sigma$ for all $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

In probability theory,

- $S = \Omega$, the sample space – the set of all possible outcomes,
- $\Sigma = \mathcal{F}$, the collection of all the events one can study.

Measures

Let (S, Σ) be a measurable space.

Definition

A *measure* is a function $\mu : \Sigma \to [0, \infty+]$ that is countably additive: if $A_n \in \Sigma$ for all $n \in \mathbb{N}$ and $A_n \cap A_m = \emptyset$ for all $n \neq m$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu\left(A_n\right)$.

Measures

Let (S, Σ) be a measurable space.

Definition

A *measure* is a function $\mu : \Sigma \to [0, \infty+]$ that is countably additive: if $A_n \in \Sigma$ for all $n \in \mathbb{N}$ and $A_n \cap A_m = \emptyset$ for all $n \neq m$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu\left(A_n\right)$.

In probability theory,

- $S = \Omega$, the sample space – the set of all possible outcomes,

Measures

Let (S, Σ) be a measurable space.

Definition

A *measure* is a function $\mu : \Sigma \to [0, \infty+]$ that is countably additive: if $A_n \in \Sigma$ for all $n \in \mathbb{N}$ and $A_n \cap A_m = \emptyset$ for all $n \neq m$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu\left(A_n\right)$.

In probability theory,

- $S = \Omega$, the sample space – the set of all possible outcomes,
- $\Sigma = \mathcal{F}$, the collection of all the events one can study,

Measures

Let (S, Σ) be a measurable space.

Definition

A *measure* is a function $\mu : \Sigma \to [0, \infty+]$ that is countably additive: if $A_n \in \Sigma$ for all $n \in \mathbb{N}$ and $A_n \cap A_m = \emptyset$ for all $n \neq m$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu\left(A_n\right)$.

In probability theory,

- $S = \Omega$, the sample space – the set of all possible outcomes,
- $\Sigma = \mathcal{F}$, the collection of all the events one can study,
- $\mu = \mathbb{P}$, the probability measure, where $\mathbb{P}(\Omega) = 1$.

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite state of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,
- $P : S \times A \times \Sigma \to [0, 1]$ is a labelled stochastic transition kernel:
  - $\forall a \in A$, $\forall s \in S$, $P(s, a, \cdot) : \Sigma \to [0, 1]$ is a probability measure,
  - $\forall a \in A$, $\forall X \in \Sigma$, $P(\cdot, a, X) : S \to [0, 1]$ is a measurable function.

Measurable Functions

Definition

Given (S, Σ), $f : S \to \mathbb{R}$ is Σ-measurable if for all $A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) \in \Sigma$.

Measurable Functions

Definition

Given (S, Σ), $f : S \to \mathbb{R}$ is Σ-measurable if for all $A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) \in \Sigma$.

Here, $P(\cdot, a, X) : S \to [0, 1]$ is a measurable function means that for all $A \in \mathcal{B}([0, 1])$, $P^{-1}(A) \in \Sigma$.

Measurable Functions

Definition

Given (S, Σ), $f : S \to \mathbb{R}$ is Σ-measurable if for all $A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) \in \Sigma$.

Here, $P(\cdot, a, X) : S \to [0, 1]$ is a measurable function means that for all $A \in \mathcal{B}([0, 1])$, $P^{-1}(A) \in \Sigma$. What is $\mathcal{B}([0, 1])$?

Measurable Functions

Definition

Given (S, Σ), $f : S \to \mathbb{R}$ is Σ-measurable if for all $A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) \in \Sigma$.

Here, $P(\cdot, a, X) : S \to [0, 1]$ is a measurable function means that for all $A \in \mathcal{B}([0, 1])$, $P^{-1}(A) \in \Sigma$. What is $\mathcal{B}([0, 1])$?

$\mathcal{B}([0, 1])$ is *the σ-algebra generated by the open sets of* $[0, 1]$.

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite state of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,
- $P : S \times A \times \Sigma \to [0, 1]$ is a labelled stochastic transition kernel:
  - $\forall a \in A$, $\forall s \in S$, $P(s, a, \cdot) : \Sigma \to [0, 1]$ is a probability measure,
  - $\forall a \in A$, $\forall X \in \Sigma$, $P(\cdot, a, X) : S \to [0, 1]$ is a measurable function.

POMDP to CMDP

Definition

A partially observable MDP (POMDP) is a tuple $(S, A, P, R, \Omega, \mathcal{O}, \gamma)$ where

- S is a (finite) set of states and A is a finite set of actions,

POMDP to CMDP

Definition

A partially observable MDP (POMDP) is a tuple $(S, A, P, R, \Omega, \mathcal{O}, \gamma)$ where

- S is a (finite) set of states and A is a finite set of actions,
- $P : S \times A \times S \to [0, 1]$ is a probabilistic transition map between states,

POMDP to CMDP

Definition

A partially observable MDP (POMDP) is a tuple $(S, A, P, R, \Omega, \mathcal{O}, \gamma)$ where

- S is a (finite) set of states and A is a finite set of actions,
- $P : S \times A \times S \to [0, 1]$ is a probabilistic transition map between states,
- $R : S \times A \to \mathbb{R}$ is a reward function,

POMDP to CMDP

Definition

A partially observable MDP (POMDP) is a tuple $(S, A, P, R, \Omega, \mathcal{O}, \gamma)$ where

- S is a (finite) set of states and A is a finite set of actions,
- $P : S \times A \times S \to [0, 1]$ is a probabilistic transition map between states,
- $R : S \times A \to \mathbb{R}$ is a reward function,
- Ω is a set of observations,

POMDP to CMDP

Definition

A partially observable MDP (POMDP) is a tuple $(S, A, P, R, \Omega, \mathcal{O}, \gamma)$ where

- S is a (finite) set of states and A is a finite set of actions,
- $P : S \times A \times S \to [0, 1]$ is a probabilistic transition map between states,
- $R : S \times A \to \mathbb{R}$ is a reward function,
- Ω is a set of observations,
- $\mathcal{O}$ is a set of conditional observation probabilities,

POMDP to CMDP

Definition

A partially observable MDP (POMDP) is a tuple $(S, A, P, R, \Omega, \mathcal{O}, \gamma)$ where

- S is a (finite) set of states and A is a finite set of actions,
- $P : S \times A \times S \to [0, 1]$ is a probabilistic transition map between states,
- $R : S \times A \to \mathbb{R}$ is a reward function,
- Ω is a set of observations,
- $\mathcal{O}$ is a set of conditional observation probabilities,
- γ is the discount factor.

POMDP to CMDP

Definition

A partially observable MDP (POMDP) is a tuple $(S, A, P, R, \Omega, \mathcal{O}, \gamma)$ where

- S is a (finite) set of states and A is a finite set of actions,
- $P : S \times A \times S \to [0, 1]$ is a probabilistic transition map between states,
- $R : S \times A \to \mathbb{R}$ is a reward function,
- Ω is a set of observations,
- $\mathcal{O}$ is a set of conditional observation probabilities,
- γ is the discount factor.

POMDP to CMDP

Definition

A partially observable MDP (POMDP) is a tuple $(S, A, P, R, \Omega, \mathcal{O}, \gamma)$ where

- S is a (finite) set of states and A is a finite set of actions,
- $P : S \times A \times S \to [0, 1]$ is a probabilistic transition map between states,
- $R : S \times A \to \mathbb{R}$ is a reward function,
- Ω is a set of observations,
- $\mathcal{O}$ is a set of conditional observation probabilities,
- γ is the discount factor.

We can represent a POMDP as a continuous MDP, where S is the simplex representing the *belief* that we are in a state in the corresponding POMDP.

Value function in CMDP

Under an optimal policy π^*, $V^*(s)$, the optimal value function is also defined via the Belllman optimality equation:

$$V^*(s) = \max_a \left(R(s,a) + \gamma \int_S P(s,a,s') V^*(s') \right) ds'$$

Sources

N. Ferns, P. Panangaden, D. Precup. *Bisimulation Metrics for Continuous Markov Decision Processes.* P.S. Castro, P.

Panangaden, D. Precup. *Equivalence Relations in Fully and Partially Observable Markov Decision Processes*