

Randomized Strategies are Useless in Markov Decision Processes

Hugo Gimbert

December 3, 2009

We show that in a Markov decision process with arbitrary payoff mapping, restricting the set of behavioral strategies from randomized to deterministic does not influence the value of the game nor the existence of almost-surely or positively winning strategies. As a corollary, we get similar results for Markov decision processes with partial observation.

1 Definitions

We use the following notations throughout the paper. Let \mathbf{S} be a countable set. The set of finite (resp. infinite) sequences on \mathbf{S} is denoted \mathbf{S}^* (resp. \mathbf{S}^ω). and \mathbf{S}^ω denotes the set of infinite sequences $u \in \mathbf{S}^\mathbb{N}$. A *probability distribution* on \mathbf{S} is a function $\delta : \mathbf{S} \rightarrow \mathbb{R}$ such that $\forall s \in \mathbf{S}, 0 \leq \delta(s) \leq 1$ and $\sum_{s \in \mathbf{S}} \delta(s) = 1$. The set of probability distributions on \mathbf{S} is denoted $\mathcal{D}(\mathbf{S})$.

Definition 1 (Markov Decision Processes). *A Markov decision process $\mathcal{M} = (\mathbf{S}, \mathbf{A}, (\mathbf{A}(s))_{s \in \mathbf{S}}, p)$ is composed of a countable set of states \mathbf{S} , a countable set of actions \mathbf{A} , for each state $s \in \mathbf{S}$, a set $\mathbf{A}(s) \subseteq \mathbf{A}$ of actions available in s , and transition probabilities $p : \mathbf{S} \times \mathbf{A} \rightarrow \mathcal{D}(\mathbf{S})$.*

In the sequel, we only consider Markov decision processes with finitely many states and actions.

An *infinite history* in \mathcal{M} is an infinite sequence in $(\mathbf{S}\mathbf{A})^\omega$. A *finite history* in \mathcal{M} is a finite sequence in $\mathbf{S}(\mathbf{A}\mathbf{S})^*$. The first state of an history is called its *source*, the last state of a finite history is called its *target*. A *strategy* in \mathcal{A} is a function $\sigma : \mathbf{S}(\mathbf{A}\mathbf{S})^* \rightarrow \mathcal{D}(\mathbf{A})$ such that for any finite history $s_0 a_1 \cdots s_n$, and every action $a \in \mathbf{A}$, $(\sigma(s_0 a_1 \cdots s_n)(a) > 0) \implies (a \in \mathbf{A}(s_n))$.

We are especially interested in strategies of the following kind.

Definition 2 (Deterministic strategies). *A strategy σ is deterministic if for every finite history h and action a , $(\sigma(h)(a) > 0) \iff (\sigma(h)(a) = 1)$.*

Given a strategy σ and an initial state $s \in \mathbf{S}$, the set of infinite histories with source s is naturally equipped with a σ -field and a probability measure denoted \mathbb{P}_s^σ . Given a finite history h and an action a , the set of infinite histories in $h(\mathbf{A}\mathbf{S})^\omega$ and $ha(\mathbf{S}\mathbf{A})^\omega$ are *cylinders* that we abusively denote h and ha . The σ -field is the one generated by cylinders and \mathbb{P}_s^σ is the unique probability measure on the set of infinite histories with source s such that for every finite history h with target t , for every action $a \in \mathbf{A}$ and for every state r ,

$$\mathbb{P}_s^\sigma(ha \mid h) = \sigma(h)(a) \ , \quad (1)$$

$$\mathbb{P}_s^\sigma(har \mid ha) = p(r|t, a) \ . \quad (2)$$

For $n \in \mathbb{N}$, we denote S_n and A_n the random variables $S_n(s_0a_1s_1 \cdots) = s_n$ and $A_n(s_0a_1s_1 \cdots) = a_n$.

Some strategies are better than other ones, this is measured by mean of a payoff function. Every Markov decision process comes with a bounded and measurable function $f : (\mathbf{S}\mathbf{A})^\omega \rightarrow \mathbb{R}$, called the *payoff function*, which associates with each infinite history h a payoff $f(h)$.

Definition 3 (Values and guaranteed values). *Let \mathcal{M} be a Markov decision process with a bounded measurable payoff function $f : (\mathbf{S}\mathbf{A})^\omega \rightarrow \mathbb{R}$. The expected payoff associated with an initial state s and a strategy σ is the expected value of f under \mathbb{P}_s^σ , denoted $\mathbb{E}_s^\sigma[f]$.*

2 Randomized strategies are useless

Randomizing his own behaviour is useless when there is no adversary to fool. This is the intuitive interpretation of the following theorem:

Theorem 4. *Let \mathcal{M} be a Markov decision process with a bounded measurable payoff function $f : (\mathbf{S}\mathbf{A})^\omega \rightarrow \mathbb{R}$, $x \in \mathbb{R}$ and s a state of \mathcal{M} . Suppose that for every deterministic strategy σ , $\mathbb{E}_s^\sigma[f] \leq x$. Then the same holds for every randomized strategy σ .*

Proof. For simplifying the notations, suppose that for every state s there are only two available actions $0, 1$ and for every action $a \in \{0, 1\}$ there are only two successor states $L(s, a)$ and $R(s, a)$ distinct and chosen with equal probability $\frac{1}{2}$.

Let σ be a strategy and s an initial state. We define a mapping

$$f_{s,\sigma} : \{L, R\}^\omega \times [0, 1]^\omega \rightarrow (\mathbf{SA})^\omega$$

that will be used for proving that \mathbb{P}_s^σ is a product measure. With every infinite word $u \in \{L, R\}^\omega$ and every sequence of real numbers $x = (x_n)_{n \in \mathbb{N}} \in [0, 1]^\omega$ between 0 and 1 we associate the unique infinite play $f_{s,\sigma}(u, x) \in (\mathbf{SA})^\omega = s_0 a_1 s_1 \cdots$ such that $s_0 = s$, for every $n \in \mathbb{N}$ if $u_n = L$ then $s_{n+1} = L(s_n, a_{n+1})$ otherwise $s_{n+1} = R(s_n, a_{n+1})$ and for every $n \in \mathbb{N}$, if $\sigma(s_0 a_1 \cdots s_n)(0) \geq x_n$ then $a_{n+1} = 0$ otherwise $a_{n+1} = 1$.

We equip $\{L, R\}^\omega$ with the σ -field generated by cylinders and the natural head/tail probability measure denoted μ_1 . We equip $[0, 1]^\omega$ with the σ -field generated by cylinders $I_0 \times I_1 \cdots \times I_n \times [0, 1]^\omega$ where I_1, I_2, \dots, I_n are intervals of $[0, 1]$, and the associated product of Lebesgue measures denoted μ_2 .

Then \mathbb{P}_s^σ is the image by $f_{s,\sigma}$ of the product of measures μ_1 and μ_2 , i.e. for every measurable set of infinite plays A ,

$$\mathbb{P}_s^\sigma(A) = (\mu_1 \times \mu_2)(f_{s,\sigma}^{-1}(A)) . \quad (3)$$

This holds for cylinders hence for every measurable A .

Now:

$$\begin{aligned} \mathbb{E}_s^\sigma[f] &= \int_{p \in (\mathbf{SA})^\omega} f(p) d\mathbb{P}_s^\sigma \\ &= \int_{(u,x) \in \{L,R\}^\omega \times [0,1]^\omega} f(f_{\sigma,s}(u, x)) d(\mu_1 \times \mu_2) \\ &= \int_{x \in [0,1]^\omega} \left(\int_{u \in \{L,R\}^\omega} f(f_{\sigma,s}(u, x)) d\mu_1 \right) d\mu_2 \end{aligned}$$

where the first equality is by definition of $\mathbb{E}_s^\sigma[f]$, the second equality is a basic property of image measures and the third equality is Fubini's theorem, that we can apply since f is bounded and the measures are probability measures.

Once x is fixed, the behaviour of strategy σ is deterministic. Formally, for every $x \in [0, 1]$ let σ_x be the deterministic strategy defined by $\sigma_x(s_0 a_1 \cdots s_n) =$

0 if and only if $\sigma(s_0 a_1 \cdots s_n)(0) \geq x_n$. Then for every $y \in]0, 1[^\omega$ and $u \in \{L, R\}^\omega$, $f_{\sigma_x, s}(u, y) = f_{\sigma, s}(u, x)$ hence:

$$\mathbb{E}_s^{\sigma_x} [f] = \int_{u \in \{L, R\}^\omega} f(f_{\sigma, s}(u, x)) d\mu_1 ,$$

and finally:

$$\mathbb{E}_s^\sigma [f] = \int_{x \in [0, 1]^\omega} \mathbb{E}_s^{\sigma_x} [f] d\mu_2 ,$$

hence the theorem, since for every x , strategy σ_x is deterministic. □

3 Applications

We provide an extension of Theorem 4 to Markov decision processes with partial observation.

A Markov decision process with partial observation is similar to a Markov decision process except every state s is labelled with a color $\text{col}(s)$ and strategies should depend only on the sequence of colors. Formally, a strategy is said to be observational if for every finite plays $s_0 \cdots s_n$ and $t_0 \cdots t_n$, if $\text{col}(s_0 \cdots s_n) = \text{col}(t_0 \cdots t_n)$ then $\sigma(s_0 \cdots s_n) = \sigma(t_0 \cdots t_n)$.

Corollary 5. *Let \mathcal{M} be a Markov decision process with a bounded measurable payoff function $f : (\mathbf{SA})^\omega \rightarrow \mathbb{R}$, $x \in \mathbb{R}$ and s a state of \mathcal{M} . Suppose that for every deterministic observational strategy σ , $\mathbb{E}_s^\sigma [f] \leq x$. Then the same holds for every randomized observational strategy σ .*

Proof. Fix an initial state s . Consider the Markov decision process whose state space is the set of finite sequences $a_0 c_0 a_1 \cdots a_n c_n \in (\mathbf{AC})^*$ of colors interleaved with actions. The initial state is the empty sequence. From state $a_0 c_0 a_1 c_1 \cdots a_n c_n$, playing action a leads to state $a_0 c_0 a_1 c_1 \cdots a_n c_n a c$ with probability:

$$\mathbb{P}_s^\sigma (A_{n+1} = a, \text{col}(S_{n+1}) = c \mid A_0 C_0 A_1 \cdots A_n C_n = a_0 c_0 a_1 \cdots a_n c_n) ,$$

and the payoff associated with an infinite play is defined by:

$$g(a_0 c_0 a_1 c_1 \cdots) = \mathbb{E}_s^\sigma [f \mid A_0 C_0 A_1 C_1 \cdots = a_0 c_0 a_1 c_1 \cdots] ,$$

where in both definitions σ is any deterministic strategy such that for every $i \in \mathbb{N}$, $\sigma(c_0 \cdots c_i) = a_{i+1}$.

The state space of this new Markov decision process is countable therefore we can apply Theorem 4 to it, which immediately gives us the result. \square