

# Equilibrium in Markov Games

Chi Li

## 1 Introduction

We begin the discussion of Markov games as an extension of repeated games.

In a repeated game, one game is played over and over. We can generalize this so that the players play a ‘random’ game (drawn from a known distribution) at each stage. If the distribution is constant, then this would just be a repeated Bayesian game. Therefore, we also want the players to have some sort of influence on the distribution depending on the current game and the actions the players take. Appealing to the definition of ‘Markov processes’, we call these games Markov games.

### Definition 1.1 (Markov Games)

A **Markov game**  $G = \{R, \mathcal{S}, \{A_r\}, \{\pi_{S,r}\}, q\}$  consists of:

1. The set of players  $R$  and the state space  $\mathcal{S}$ .
2. For each state  $S \in \mathcal{S}$  and each player  $r \in R$ , a action (strategy) set  $A_r(S)$  for the player in the game  $S$ .
3. For each state  $S \in \mathcal{S}$  and each player  $r \in R$ , a payoff function  $\pi_{S,r} : \prod_{r' \in R} A_{r'}(S) \rightarrow \mathbb{R}$  for player  $r$  in game  $S$ .
4. A transition function  $q$  for the next state given by  $q(S^{t+1}|S^t, a^t)$  where  $S^t$  is the current state at time  $t$  and  $a^t$  is the set of actions chosen by each player.

If required, we can also introduce a discount rate  $\delta$ .

**Remark.** Under this definition, we can treat a repeated game as a Markov game with one state.

Our first question in characterizing Markov game equilibria is as follows: is there an equilibrium where the decisions of each player are also ‘Markov’ (in the spirit of the game)? That is, is there an equilibrium where each players strategy is only determined by the current state?

### Definition 1.2 (Markov Perfect Equilibria)

A **Markov strategy** for player  $r$  is  $\Theta : \mathcal{S} \rightarrow \sqcup_{S \in \mathcal{S}} A_r(S)$  that determines the action at time  $t$


$$a_r^t = \Theta(S^t).$$

A **Markov Perfect Equilibrium** is a Nash equilibrium such that each player’s strategy is Markov.

**Remark.** Assume the payoffs are uniformly bounded and the number of states are finite (or at least discrete). Then the discounted payoffs for each player exist in limit almost surely and converge in expectation. Thus, if all players are using Markov strategies, we can solve for the expected discounted payoffs by writing recurrence relations for payoffs for each state and solving the linear system. In fact, this is the usually approach for dealing Markov games, as we will see in the examples later.

### Proposition 1.3


Markov strategies are a best response to a Markov strategy.

*Proof.* Suppose everyone except player  $r$  is using a Markov strategy. Then if state  $S$  is reached twice in the game at times  $t_1$  and  $t_2$ , the subgames starting from the two times are indistinguishable for player  $r$  except possibly a discounted factor in payoff. Thus a best response for  $t_1$  is also a best response for  $t_2$ . Thus there is a best response independent of play history. 

We are now prepared to state the existence of Markov Perfect Equilibria.

### Theorem 1.4 (Existence of Markov Perfect Equilibria [3])

Let  $G$  be a finite Markov game. That is  $|\sqcup_{r \in R, S \in \mathcal{S}} A_r(S)| < \infty$ . Then  $G$  has a mixed-strategy Markov perfect equilibrium.

*Proof.* We consider a (one-stage) game  $G'$ , where the strategy set of each player correspond to a Markov strategy and the payoffs are the expected values of the payoffs in  $G$  with the corresponding Markov strategies. By Nash, this game  $G'$  has a mixed Nash equilibrium. We want to show that the corresponding mixed Markov strategies define a Markov Perfect equilibrium. This is true as by Proposition 1.3, each player is best responding to all other players. 

As a simple example, an infinitely repeated game that has a Nash equilibrium. If the equilibrium is unique, then the only Markov Perfect equilibrium is that all players use the same Nash Equilibrium strategy at every stage.

## Monopoly as a Markov Game

We model the board game Monopoly as a Markov game. Suppose there are  $N$  players in the game. We can represent each state of the board by the following parameters: The location, assets (cash and properties) of each player and the player making the decision in that state.

Upon arriving each state, the player can choose to buy the property, or pay rent. The player making the decision can also propose a trade with other players, which is essentially a proposal to change the state to another with the same sum of possessed assets and the same location. More rules can be incorporated to increase the complexity of the game, but these are still finite (going to jail, chance, passing go etc).

We can model the transitions as follows: The next player making the decision advances his position according to the value determined by the sum of dice rolls. This property is Markov. The payoff at each stage is 0, as nobody wins yet. The only states with payoffs are the ending states when all but one player are bankrupt. The winning player gets a payoff of 1 and the rest gets 0. The game also terminates at this stage.

Unfortunately, Monopoly is not a finite game. This is because the bank never runs out of money, so the assets parameter can take infinitely many values. Other parameters are bounded - specifically

the number of houses and hotels are capped. A way to mitigate this is to look for  $\epsilon$ -stable Nash equilibria, by introducing a future discount factor  $\delta$ . For sufficiently wealthy players, it would take too many turns drain their money, so the discounted payoff for even the winning player is  $< \epsilon$ . Thus, we can limit the wealth of each player to be  $\leq M$  for some absurd amount of wealth  $M$  depending on how patient the players are. Under this new modified game, there is a Markov perfect equilibrium. Moreover, since players do not make simultaneous decisions and have perfect information, the equilibrium is pure.

## 2 Case study - Markov Perfect Equilibria of a bargaining game

We study a one-player Markov game (and will generalize it to two players), both of which were studied by Cripps [1], which seems to be inspired by stock market prices.

### Example 2.1 (Optimal stopping)

Suppose Bob has an item whose value on day  $d$  depends on a Markov process with countable states i.e. there is some function  $f : \mathcal{S} \rightarrow \mathbb{R}$   $V_d = f(S_d) \geq 0$ . We number these states  $1, 2, \dots$  such that  $1 = f(1) > f(2) > \dots$ . Without loss of generality, we can set that the infimum of  $f$  is 0. On any day  $d$ , Bob can choose to use the item with payoff  $\delta^d V_d$  and the game terminates. Thus we can model the transition probability to be independent of Bob's action and write the transition probability from state  $n$  to  $m$  as  $q(m|n)$ .

There is an intuitive 'best strategy'. On day  $i$ , Bob can use the item for  $f(i)$ , or wait a day and repeat the strategy. Thus, the optimal strategy of stopping time  $\tau$  can be identified with a subset  $G \subseteq \mathbb{N}$ , such that the  $\tau$  is the first time the state enters the set  $G$ . We take the supremum of this payoff over all subsets  $G$ . Let  $v(i)$  describe the supremum of payoff across all strategies with  $S^1 = i$ . We thus have

$$v(i) = \max(f(i), \delta \sum_j p(j, i) v(j)) \geq f(i).$$

This means we must have the optimal stopping criterions given by  $G = i : v(i) = f(i)$ .

### Lemma 2.2 (Cripps)

If the Markov Chain is also irreducible and aperiodic, then the solution  $\tau^*$  described above gives a Markov Perfect equilibria. That is, the solution produces the expected payout

$$v(i) = E_i \delta^{\tau^*} f(S_{\tau^*}).$$

### Example 2.3 (Markov Bargaining)

We now extend this game to two players. We use the same item as in example 2.1. Now Alice owns the item and wants to sell it to Bob. On each day, the value of the item is known to both players. Alice can propose to sell the item for  $x_1$ , to which Bob can accept or reject. If Bob rejects, the game continues into the next day and Bob proposes a bid. They alternate bargains until Bob decides to buy. If Bob accepts the offer on day  $d$ , Alice gets a payoff of  $x_d$ . Upon obtaining the item, Bob can use it any time in the future according to the optimal stopping strategy with the same payoff as in the previous example. Furthermore, Alice and

Bob have discount factors  $\gamma$  and  $\delta$  respectively. The payoffs for Alice and Bob are thus


$$(\gamma^d x_d, \delta^d (v(i) - x_d)).$$

#### Theorem 2.4

There exists a Markov Perfect equilibrium for the aperiodic irreducible Markov Bargaining game.


*Sketch of proof.* Similar to the infinite horizon bargaining game, we assume that the payoffs for Alice and Bob at each state  $i$  are  $\alpha(i)$  and  $\beta(i)$  respectively. Recursively, if both players are best responding to each other we must have the bidder proposing the payoffs  $(v(i) - \beta(i), \beta(i))$ , and the seller proposing the payoffs  $(v(i) - \delta E[\beta(S^2)|S^1 = i], \delta E[\beta(S^2)|S^1 = i])$ . Each player must wait two turns to repropose, so we need to solve

$$\begin{aligned}\alpha(i) &= \max(v(i) - \delta E[\beta(S^2)|S^1 = i], \gamma^2 E[\alpha(S^3)|S^1 = i]), \\ \beta(i) &= \max(v(i) - \gamma E[\alpha(S^3)|S^2 = i], \delta^2 E[\beta(S^4)|S^2 = i])\end{aligned}$$

The remaining parts of the proof boils down to constructing the solution to this system of equations. 

#### Lemma 2.5

Let  $v_\delta$  and  $v_\gamma$  denote the expected payoffs in the Optimal Stopping game with discount factor  $\delta$  and  $\gamma$  respectively. Then if  $v_\delta(i) = v_\gamma(i) = f(i)$  for some state  $i$ , the bargaining will result in an agreement in state  $i$ .

*Sketch of proof.* We prove it for Alice. Suppose Alice does not want an agreement. We have  $\alpha(i) \geq v_\delta(i) - \delta E[\beta(S^2)|S^1 = i] \implies \alpha(i) + \delta E[\beta(S^2)|S^1 = i] \geq v_\delta(i)$ . Notice that left hand side describes the sum of discounted payoffs when agreement is reached, possibly in the future. Recall that when an agreement is reached, the sum of their payoffs is exactly  $v_\delta(S^d)$  before discounting. But for any (future) stopping time, the expected value of  $\max(\delta, \gamma)^\tau S^\tau$  is less than  $v_\delta(i)$  by the condition  $v_\delta(i) = v_\gamma(i) = f(i)$ . So we must also have equality. 

#### Theorem 2.6 (Cripps)

With no additional conditions on the Markov process, if  $\delta \geq \gamma$  and  $\gamma < 1$ , then there is a unique Markov Equilibrium characterized by the equations

$$\begin{aligned}\alpha(i) - \gamma \delta E[\alpha(S^3)|S^1 = i] &= v_\delta(i) - \delta E[v_\delta(S^2)|S^1 = i] \\ \beta(i) - \gamma \delta E[\beta(S^3)|S^1 = i] &= v_\delta(i) - \gamma E[v_\delta(S^2)|S^1 = i].\end{aligned}$$

The proof is very similar to solving the equations Theorem 2.4, except that it deals with the supremum and infimum of all possible Nash equilibrium payoffs.

### 3 Justification and implications of Markov Perfect Equilibria

Using the Markov Perfect equilibria of Monopoly, we can see that trading is not useful for winning when there are only two players. Assigning a probability of winning for each state, the only way a trade reaches agreement is that both players increase in winning probability, which is impossible. When there are three or more players, it is possible that all players involved benefit at the expense of the players outside the trade, such as two people exchanging to complete a set of properties. Trading in two players is expected not because that it is beneficial for either player in the sense of winning probability, but the game is expedited so the discounted payoffs are increases. In particular, if players are not equally patient, the more patient player usually benefits from the trades when considering winning probability.

The unique equilibrium in the bargaining example is qualitatively similar to the deterministic bargaining problem solved by Rubinstein [1], with discounted payoffs based on the transition multiplied by the discount factor, and  $v_\delta$  describing the absolute value of the item. However, the proof requires the seller to be less patient than the buyer. In this equilibrium, the agreement is reached before the stopping time for using is reached. That is, Bob never finds himself in a situation where he would want to use the item but does not own it. Because Bob ends up with the item, it is never beneficial for the buyer to delay agreement. The same equilibrium cannot be used for when the seller is more patient than the buyer, as the seller may find it beneficial to delay agreements. Consider the following:

#### Example 3.1 (Delaying in agreement for seller)

Let  $f(S^1) = 1/2$  and  $f(S^d) = 1$  onwards, and  $\gamma > \delta$ . If  $\delta < 1/2$ , the Bob will use the item on day 1 if it is bought. From day 2 onwards the value of the item is deterministic, so the Rubenstein's solution applies with payoffs in day 2 equalling

$$(\gamma(1 - \delta)/(1 - \delta\gamma), (1 - \gamma)/(1 - \delta\gamma)).$$

So in day 1, Alica can either wait until day 2 to get payoff of  $\gamma^2(1 - \delta)/(1 - \delta\gamma)$ , or offer based on the discounted payoff  $(1/2 - \delta(1 - \gamma)/(1 - \delta\gamma), \delta(1 - \gamma)/(1 - \delta\gamma))$ . Take  $\gamma = 1$ ,  $\delta = 0.000001$ . Even Bob wants to use the item on day 1, the agreement is reached on day 2.

Markov Perfect Equilibria are also inherently non-cooperative. This is useful for modelling the stock market with too many people to reach a consensus. For smaller settings, cooperative solutions can be reached, such as collusion. There is an analogous folk theorem for Markov Games. I shall state this but not prove, as it takes too long to set up (which will be evident in its statement). Moreover, the enforcing strategizing is somewhat similar in spirit to the original folk theorem, that is, through the punishment regime.

#### Theorem 3.2 (Folk Theorem for Markov Games [2])

Let  $G$  be a Markov Game with discounted payoff factor  $\delta < 1$ . Assume that for any set of strategies the long-term payoff and long term min-max of agents is independent of starting state. We call a outcome **feasible** if it lies in the convex hull of the discounted payoffs over all pure Markov strategies. We call a outcome **individually rational** if each agent's long term payoff is at least the long term min-max payoff of the agent.

Assume that the set of feasible payoffs (the convex hull) has dimension equal to the number

of players. Then for  $\epsilon > 0$  and any feasible and individually rational outcome  $\pi$  there is  $\tilde{\delta} < 1$  such that for any  $\delta \geq \tilde{\delta}$  there is an equilibrium strategy such that the long term payoffs converge to within  $\epsilon$  of  $\pi$ .

## References

- [1] Martin W. Cripps. “Markov bargaining games”. In: *Journal of Economic Dynamics and Control* 22.3 (1998), pp. 341–355. ISSN: 0165-1889. DOI: [https://doi.org/10.1016/S0165-1889\(97\)00059-6](https://doi.org/10.1016/S0165-1889(97)00059-6). URL: <https://www.sciencedirect.com/science/article/pii/S0165188997000596>.
- [2] Prajit K. Dutta. “A Folk Theorem for Stochastic Games”. In: *Journal of Economic Theory* 66.1 (1995), pp. 1–32. ISSN: 0022-0531. DOI: <https://doi.org/10.1006/jeth.1995.1030>. URL: <https://www.sciencedirect.com/science/article/pii/S0022053185710307>.
- [3] Drew Fudenberg and Jean Tirole. *Game Theory*. The MIT Press, 1991.