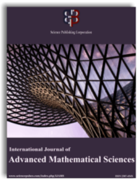
International Journal of Advanced Mathematical Sciences, 11 (2) (2025) 17-24

International Journal of Advanced Mathematical Sciences

Website: www.sciencepubco.com/index.php/IJAMS
doi: https://doi.org/10.14419/etx28e49
Research paper



Nonzero-Sum Markov Games with Finite Random Horizon

R. Israel Ortega-Gutiérrez[1], Hugo Cruz-Suárez[2*] and Adriana González-Quiroz[3]

[1]*Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y Río Verde, Col. San Manuel, CU, Puebla, Pue. 72570. México.*
[2]*Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y Río Verde, Col. San Manuel, CU, Puebla, Pue. 72570. México.*
[3]*Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y Río Verde, Col. San Manuel, CU, Puebla, Pue. 72570. México.*
[*]*Corresponding author E-mail:hugo.cruz@correo.buap.mx*

Abstract

This manuscript aims to establish the existence of a Nash equilibrium in nonzero-sum Markov games with a random horizon of finite support. The proof relies on dynamic programming techniques adapted to the stochastic nature of the horizon and the interaction between players. Introducing a random horizon with finite support allows for a more realistic modeling of scenarios in which the duration of the game is uncertain and influenced by exogenous random events. To illustrate the applicability of the theoretical results, we examine a dynamic game version of the Great Fish War, which models competition over a renewable resource under uncertainty about the duration of exploitation. This framework enhances the applicability of Markov game theory to decision-making contexts where time horizons are unpredictable.

Keywords: *Dynamic programming, Nonzero-Sum stochastic games, Random horizon*

1. Introduction

This article focuses on nonzero-sum Markov games with a random horizon. Markov games originate from the foundational ideas introduced by Lloyd Shapley [18], who developed stochastic games in both finite and infinite horizon settings, thereby introducing a class of dynamic games in which the transition to the next state depends solely on the current state and the actions taken by the players, a structure that is now formally recognized as the Markov property. On the other hand, the notion of a random horizon was first explored by Levhari in [11], in the context of Markov decision processes (MDPs). This naturally motivates the extension of this concept to the framework of Markov games.

The importance of considering a random horizon lies in the need to model the duration of strategic interactions in a more realistic manner. While game-theoretic models traditionally assume either a finite or infinite horizon [14], in many real-world scenarios the actual length of the game is uncertain and driven by unpredictable external factors. For example, if two firms compete in a market by offering the same product, the sudden bankruptcy of one of them may abruptly terminate the competitive process, leading to an unanticipated decision horizon. Moreover, random horizons arise naturally in a variety of domains where the time of termination is inherently uncertain: in finance, it may depend on a sudden asset collapse; in medicine, on a patient's recovery time; and in natural resource management, on the survival of a population.

The motivation for this line of research arises from the limited literature on dynamic Markov games with random horizons. Among the few contributions in this area are [7], which analyzes discrete-time zero-sum Markov games with random horizons, and [8], which studies nonzero-sum Markov games under a probabilistic criterion, allowing both the transition probabilities and the reward functions to vary over time. This gap is significant, since nonzero-sum games capture a broader spectrum of real-world interactions where players' interests may be partially opposed, thereby modeling scenarios that involve cooperation, competition, or the coexistence of individual goals. Furthermore, the theory of random-horizon MDPs developed in [4] provides a foundational framework that supports the extension of these ideas to strategic multi-agent settings. Related strands of research include optimal stopping problems that explicitly consider randomness in the termination of an MDP, as studied in [5, 19], as well as recent work on random-horizon MDPs such as [1, 6], which investigate risk-sensitive approaches and regret bounds in continuous-time contexts.

In this context, we study a two-player nonzero-sum Markov game in which the players act rationally, seeking to optimize their cumulative payoffs. To analyze the strategies adopted throughout the game, we introduce an optimality criterion that evaluates the effectiveness of each strategy according to the structure and objectives of the game. This criterion serves as the basis for defining a Nash equilibrium, which identifies mutually optimal strategies for the players involved. The existence and characterization of such equilibria are established through a dynamic programming approach adapted to the random-horizon setting, allowing for a recursive formulation of the value functions and equilibrium strategies.

To illustrate the relevance and applicability of the Nash equilibrium in this setting, we examine the Great Fish War as a representative example. The Great Fish War has gained prominence due to its ability to model international conflicts over shared fishery resources. These disputes typically arise when multiple agents compete for limited aquatic resources, leading to tensions over access and sustainability. Documented examples include the Turbot War between Canada and Spain (1995) [15], the Pacific Salmon War in the 1990s [13], and ongoing conflicts in Yucatán, Mexico [3] related to unregulated octopus fishing. The original model, introduced by Levhari and Mirman [12] in the context of Markov decision processes, was later extended to non-cooperative nonzero-sum games [10, 16], establishing conditions for the existence of Nash equilibria. More recent contributions have generalized the model to aquaculture settings, incorporating utility maximization under fishery production constraints [2, 17].

The manuscript is structured as follows. We begin by presenting the theoretical framework and, using dynamic programming techniques, prove the existence of a Nash equilibrium in nonzero-sum Markov games where each player faces an independent random horizon with finite support. The applicability of the results is then illustrated through the analysis of the Great Fish War, a dynamic game modeling strategic resource exploitation under uncertainty.

2. Non-zero-sum Markov games with random horizon

This section presents the formal model of nonzero-sum Markov games with random horizons. The analysis is restricted to the two-player case. Unless otherwise specified, references to player i will correspond to $i = 1,2$.

Definition 1. A *nonzero-sum Markov game* with random horizon is defined by the tuple

$$GM := \left\{X, (A_i, A_i(x), r_i, \tau_i)_{i=1,2}, Q\right\}.$$

where:

(a) X is a Borel space, representing the state space.

(b) For each player $i = 1,2$:

   (i) A_i is a Borel space denoting the set of actions.

   (ii) $A_i(x) \subseteq A$ is the set of admissible actions at state $x \in X$. The joint admissible action set at state $x \in X$ is given by $A(x) := A_1(x) \times A_2(x)$, and the set of admissible state-action triples is

$$\mathbb{K} := \{(x, a_1, a_2) | x \in X, (a_1, a_2) \in A(x)\}.$$

   (iii) $r_i : \mathbb{K} \to \mathbb{R}$ is a measurable stage reward function.

   (iv) τ_i is a discrete random variable defined on a probability space $(\Omega_i, \mathscr{F}_i, \mathbb{P}_i)$ with known distribution $\mathbb{P}_i(\tau_i = t) := \rho_{i,t}$, $t = 0, 1, 2, \ldots, T_i$, where $T_i \in \mathbb{N}$ and the variables τ_1 and τ_2 are assumed to be independent.

(c) Q is a stochastic kernel on X given $\mathbb{K}$, which represents the transition law of the system.

Remark 1. Observe that in Definition 1, it is considered that each player is assigned an individual random horizon τ_i, allowing for asymmetric termination of the game. The game is played as a two-player interaction while both players remain active. Once one player reaches their horizon, the interaction ends, and the remaining player, if still within their horizon, continues in a single-agent setting that corresponds to a Markov decision process [9]. This formulation naturally includes the case in which both horizons coincide.

The game proceeds as follows. At each time $t \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, the players observe the current state $x_t \in X$ and simultaneously select their actions $(a_{1,t}, a_{2,t}) \in A(x_t)$, independently of one another. Each player then receives a stage reward $r_i(x_t, a_{1,t}, a_{2,t})$. Subsequently, the state of the system evolves according to the transition kernel $Q(\cdot | x_t, a_{1,t}, a_{2,t})$, and this process is repeated at each stage until the game ends based on the random horizons assigned to the players.

Definition 2. Let $H_0 := X$, and for each $t \in \mathbb{N}$, define $H_t := \mathbb{K} \times H_{t-1}$. An element $h_t \in H_t$ is given by

$$h_t = (x_0, a_{1,0}, a_{2,0}, ..., x_{t-1}, a_{1,t-1}, a_{2,t-1}, x_t)$$

and represents the *history of the game up to time t*, where $(x_j, a_{1,j}, a_{2,j}) \in \mathbb{K}$ for each $j = 0, 1, ..., t-1$ and $x_t \in X$.

Definition 3. A *strategy* for player i is defined as a sequence of stochastic kernels $\pi_i = \{\pi_{i,t}\}_{t \in \mathbb{N}_0}$, where each $\pi_{i,t}$ assigns to every history $h_t \in H_t$ a probability distribution supported on the set of admissible actions, i.e. $\pi_{i,t}(A_i(x_t)|h_t) = 1$, $\forall h_t \in H_t$, and $t \in \mathbb{N}_0$. The collection Π_i denotes the set of all strategies for player i. A pair $\pi = (\pi_1, \pi_2) \in \Pi := \Pi_1 \times \Pi_2$ is called a *strategy profile*.

To describe the structure of the strategies that arise in equilibrium, we introduce a subclass of admissible strategies known as deterministic Markov strategies. These depend only on the current state and time, and not on the full history of the game.

Definition 4. Let $\mathbb{F}_i$ be the set of all measurable functions $f : X \to A_i$ such that $f(x) \in A_i(x)$ for all $x \in X$. A strategy $\pi_i = \{\pi_{i,t}\} \subset \Pi_i$ is called a *deterministic Markov strategy* if there exists a sequence $\{f_{i,t}\}$ with $f_{i,t} \in \mathbb{F}_i$ such that, for every $h_t \in H_t$, the distribution $\pi_{i,t}(\cdot|h_t)$ is the Dirac measure concentrated at $f_{i,t}(x_t) \in A_i(x_t)$.

Now, with the goal of evaluating the system's response to each strategy profile, we introduce an optimality criterion. To this end, it is necessary to construct a unified probability space that jointly describes both the evolution of the system and the termination times. For this purpose, we extend the canonical measurable space $(\Omega', \mathscr{F}')$, which supports the stochastic process $\{(x_t, a_{1,t}, a_{2,t})\}_{t\geq 0}$, by incorporating the random variables τ_1 and τ_2, associated with each player's horizon. The resulting space is denoted by $(\Omega, \mathscr{F}) := (\Omega' \times \Omega_1 \times \Omega_2, \mathscr{F}' \otimes \mathscr{F}_1 \otimes \mathscr{F}_2)$, equipped with the product measure $\mathbb{P} := \mathbb{P}_x^{\pi_1,\pi_2} \otimes \mathbb{P}_1 \otimes \mathbb{P}_2$. Here $\mathbb{P}_x^{\pi_1,\pi_2}$ governs the evolution of the state-action process: if the process starts at the initial state $x_0 = x$ and players 1 and 2 follow the strategies π_1 and π_2, then this probability law induces a distribution over all possible trajectories $(x_0, a_{1,0}, a_{2,0}, x_1, a_{1,1}, a_{2,1}, x_2, \dots)$. In this way, $\mathbb{P}_x^{\pi_1,\pi_2}$ describes how the possible trajectories of the game are distributed under a given strategy profile, while $\mathbb{P}_i$ denotes the probability law of τ_i for each player i. The validity of this construction follows from the Theorem of C. Ionescu-Tulcea [9]. The stochastic process $\{x_t\}$ defined on $(\Omega, \mathscr{F}, \mathbb{P})$ is called the state process of the game.

Definition 5. Let $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ be a pair of strategies and let $x \in X$ be the initial state. For each player $i \in \{1,2\}$, the *total expected discounted payoff with random horizon* τ_i is defined as

$$\mathscr{J}_{i,\tau_i}(x, \pi_1, \pi_2) := \mathbb{E}\left[\sum_{t=0}^{\tau_i} \beta_i^t r_i(x_t, a_{1,t}, a_{2,t})\right], \tag{1}$$

where: $\beta_i \in (0,1)$ is the discount factor of player i, and $\mathbb{E}$ denotes the expectation taken with respect to the product measure $\mathbb{P}$.

Having introduced the performance criterion that assigns to each player their expected total discounted reward up to a random termination time, we now proceed to define the solution concept adopted in this framework. In the context of nonzero-sum Markov games, a natural and widely accepted notion of solution is the Nash equilibrium [10], which ensures that no player can unilaterally improve their expected outcome by deviating from their prescribed strategy.

Definition 6. A strategy profile $(\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ is called a *Nash equilibrium* for the Markov game with random horizon if, for every initial state $x \in X$, the following conditions hold:

$$\mathscr{J}_{1,\tau_1}(x, \pi_1^*, \pi_2^*) \geq \mathscr{J}_{1,\tau_1}(x, \pi_1, \pi_2^*) \quad \text{for all } \pi_1 \in \Pi_1,$$
$$\mathscr{J}_{2,\tau_2}(x, \pi_1^*, \pi_2^*) \geq \mathscr{J}_{2,\tau_2}(x, \pi_1^*, \pi_2) \quad \text{for all } \pi_2 \in \Pi_2.$$

Given such an equilibrium, the *value function* for player $i \in \{1,2\}$ at state $x \in X$ is defined by

$$\mathscr{V}_{i,\tau_i}^*(x) := \mathscr{J}_{i,\tau_i}(x, \pi_1^*, \pi_2^*),$$

which represents the expected total discounted reward that player i obtains under the equilibrium strategy profile.

Remark 2. An equivalent characterization of the Nash equilibrium can be expressed in terms of the players' value functions. A strategy profile $(\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ is a Nash equilibrium if and only if, for every initial state $x \in X$,

$$\mathscr{V}_{1,\tau_1}^*(x) := \mathscr{J}_{1,\tau_1}(x, \pi_1^*, \pi_2^*) = \sup_{\pi_1 \in \Pi_1} \mathscr{J}_{1,\tau_1}(x, \pi_1, \pi_2^*),$$
$$\mathscr{V}_{2,\tau_2}^*(x) := \mathscr{J}_{2,\tau_2}(x, \pi_1^*, \pi_2^*) = \sup_{\pi_2 \in \Pi_2} \mathscr{J}_{2,\tau_2}(x, \pi_1^*, \pi_2).$$

The previous definitions establish the optimality criterion and the notion of equilibrium in the context of Markov games. In what follows, we develop a dynamic programming approach that allows the characterization and computation of equilibrium strategies under suitable structural assumptions.

3. Nash Equilibria in Random-Horizon Markov Games

Consider the nonzero-sum Markov game model with random horizon, described in Definition 1: $GM = \{X, (A_i, A_i(x), r_i, \tau_i)_{i=1,2}, Q\}$. The strategic behavior of each player is evaluated according to the expected total discounted reward with random horizon, as defined in equation (1). This section develops a dynamic programming formulation to characterize equilibrium strategies and establish existence results. To establish the dynamic programming framework we begin by stating the structural assumptions on the game components.

Assumption 1. For each initial state $x \in X$ and each strategy profile $\pi \in \Pi$, the stochastic process $\{(x_t, a_{1,t}, a_{2,t}) | t = 0,1,2,\dots\}$ induced by π is independent of the random horizons τ_1 and τ_2.

Remark 1. Assumption 1 is natural in many stochastic models. For instance, in reliability and survival analysis, the dynamics of the underlying process (e.g., the operation of a machine or the evolution of a biological system) are often modeled independently of the random horizon, which is determined by an exogenous lifetime or failure time. Similarly, in economics and operations research, the termination of a decision process may be governed by external factors (such as contract expiration or budget constraints) that do not influence the state transitions of the system itself.

The proof of Lemma 1 follows the same reasoning as that presented in Remark 3.1 of [4], and is included here for completeness.

Lemma 1. Under Assumption 1, the expected total discounted reward with random horizon for player i can be expressed as:

$$\mathscr{J}_{i,\tau_i}(x, \pi_1, \pi_2) = \mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{T_i} P_{i,t} \beta_i^t r_i\left(x_t, a_{1,t}, a_{2,t}\right)\right] \tag{2}$$

for $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2, x \in X$, where

$$P_{i,t} := \sum_{n=t}^{T_i} \rho_{i,n} = \mathbb{P}_i(\tau_i \geq t), t = 0,1,2,\dots,T < \infty.$$

Here, the expectation $\mathbb{E}_x^{\pi_1,\pi_2}$ is taken with respect to the probability measure $\mathbb{P}_x^{\pi_1,\pi_2}$ on the canonical space $(\Omega', \mathscr{F}')$.

Proof. Let $(\pi_1,\pi_2)\in\Pi_1\times\Pi_2$ and $x\in X$. From the definition of the expected total discounted reward with random horizon for player i, we have

$$\mathscr{J}_{i,\tau_i}(x,\pi_1,\pi_2)=\mathbb{E}\left[\sum_{t=0}^{\tau_i}\beta_i^t r_i(x_t,a_{1,t},a_{2,t})\right].$$

Since τ_i is a discrete random variable with finite support $\{0,1,\ldots,T_i\}$, we can use the indicator function $\mathbb{I}_{\{t\leq\tau_i\}}$ to rewrite the sum as

$$\mathscr{J}_{i,\tau_i}(x,\pi_1,\pi_2)=\mathbb{E}\left[\sum_{t=0}^{T_i}\mathbb{I}_{\{t\leq\tau_i\}}\beta_i^t r_i(x_t,a_{1,t},a_{2,t})\right].$$

By linearity of expectation and Assumption 1, we obtain

$$\begin{aligned}\mathscr{J}_{i,\tau_i}(x,\pi_1,\pi_2)&=\sum_{t=0}^{T_i}\beta_i^t\cdot\mathbb{E}_x^{\pi_1,\pi_2}\left[r_i(x_t,a_{1,t},a_{2,t})\right]\mathbb{P}_i(\tau_i\geq t)\\&=\mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{T_i}P_{i,t}\cdot\beta_i^t r_i(x_t,a_{1,t},a_{2,t})\right],\end{aligned}$$

which completes the proof. □

We now state additional assumptions to ensure the existence of a Nash equilibrium.

Assumption 2. For each player $i=1,2$ and every state $x\in X$, the following conditions hold:

(a) The action set $A_i(x)$ is compact.
(b) The stage reward function $r_i(x,a_1,a_2)$ is upper semicontinuous in $(a_1,a_2)\in A_1(x)\times A_2(x)$.
(c) The transition kernel $Q(\cdot|x,a_1,a_2)$ is weakly continuous in $(a_1,a_2)\in A_1(x)\times A_2(x)$.

Theorem 1. Let $J_{i,0},J_{i,1},...,J_{i,T_i+1}$ be a sequence of real-valued functions on X defined recursively for each player $i=1,2$ as follows:

$$J_{i,T_i+1}(x):=0,\qquad x\in X$$

and for each $t=T_i,T_i-1,...,0,$

$$J_{i,t}(x):=\max_{a_i\in A_i(x)}\left[P_{i,t}\beta_i^t r_i(x,a_1,a_2)+\int_X J_{i,t+1}(y)Q(dy|x,a_1,a_2)\right],x\in X. \tag{3}$$

Under Assumption 1 and Assumption 2, the following hold:

a) The functions $J_{i,t}$ are measurable for all t, and for each $t=0,1,...,T_i$, there exists a measurable selector $f_{i,t}^*\in\mathbb{F}_i$ such that $f_{i,t}^*(x)\in A_i(x)$ achieves the maximum in (3) for all $x\in X$, that is,

$$J_{i,t}(x)=P_{i,t}\beta_i^t r_i(x,f_{1,t}^*(x),f_{2,t}^*(x))+\int_X J_{i,t+1}(y)Q(dy|x,f_{1,t}^*(x),f_{2,t}^*(x)).$$

b) Let $\pi^*=(\pi_1^*,\pi_2^*)\in\Pi_1\times\Pi_2$, with $\pi_i^*=\{f_{i,0}^*,f_{i,1}^*,...,f_{i,T}^*\}$. Then, for every $x\in X$,

$$\mathscr{J}_{i,\tau_i}(x,\pi_1^*,\pi_2^*)=\mathscr{V}_{i,\tau_i}^*(x)=J_{i,0}(x),$$

and the strategy profile π^* is a Nash equilibrium.

Proof. The proof proceeds by first analyzing the problem from the perspective of player 1, assuming that both players act rationally. That is, player 2 is assumed to follow the strategy π_2^*, and player 1 optimizes accordingly. Note that, under Assumption 2, the existence of the strategies π_1^* and π_2^* is guaranteed. Indeed, for each player i, the dynamic programming recursion given by equation (3) involves the maximization of a function that is upper semicontinuous in the action variables and defined over compact-valued, measurable multifunctions $A_i(x)\subset A_i$. Therefore, by standard results on measurable selection (see, e.g., Section 3.3 in [9]), there exists a finite sequence of measurable selectors $\{f_{i,t}^*\}_{t=0}^{T_i}$, each attaining the maximum in equation (3). These selectors define an admissible deterministic Markov strategy $\pi_i^*\in\Pi_i$. Let $\pi_1=\{a_{1,t}\}_{t=0}^{T_1}\in\Pi_1$ be an arbitrary strategy for player 1, and let $\pi_2^*=\{f_{2,t}^*\}_{t=0}^{T_2}\subset\mathbb{F}_2$ be fixed. For each $t=0,1,\ldots,T_1$, define

$$R_{1,t}(x,\pi_1,\pi_2^*):=\mathbb{E}_x^{\pi_1,\pi_2^*}\left[\sum_{n=t}^{T_1}P_{1,n}\beta_1^n r_1(x_n,a_{1,n},f_{2,n}^*)\right],$$

and set $R_{1,T_1+1}(x,\pi_1,\pi_2^*):=0$. Note that $R_{1,t}$ represents the expected cumulative reward obtained by player 1 from stage t onward. We will prove the following two statements:

1) $R_{1,t}(x,\pi_1,\pi_2^*)\leq J_{1,t}(x)$ for all $t=0,\ldots,T_1+1$.
2) If $\pi_1=\pi_1^*$, then $R_{1,t}(x,\pi_1^*,\pi_2^*)=J_{1,t}(x)$.

1) The proof will proceed by backward induction on t.
Base case. For $t = T_1 + 1$, by definition of the functions $R_{1,t}$ and $J_{1,t}$, we have

$$R_{1,T_1+1}(x,\pi_1,\pi_2^*) = 0 = J_{1,T_1+1}(x).$$

Inductive step. Suppose that, for some $t \in \{0,\dots,T_1\}$,

$$R_{1,t+1}(x,\pi_1,\pi_2^*) \le J_{1,t+1}(x), \quad \forall x \in X.$$

We now prove that the inequality also holds at stage t. Using the definition of $R_{1,t}$, we obtain:

$$\begin{aligned}
R_{1,t}(x,\pi_1,\pi_2^*) &= \mathbb{E}^{\pi_1,\pi_2^*}\left[\sum_{n=t}^{T_1} P_{1,n}\beta_1^n r_1(x_n,a_{1,n},f_{2,n}^*)\Big| X_t = x\right] \\
&= \mathbb{E}^{\pi_1,\pi_2^*}\left[P_{1,t}\beta_1^t r_1(x_t,a_{1,t},f_{2,t}^*) + \sum_{n=t+1}^{T_1} P_{1,t}\beta_1^n r_1(x_n,a_{1,n},f_{2,n}^*)\Big| X_t = x\right] \\
&= \int_{A_1}\left[P_{1,t}\beta_1^t r_1(x,a_1,f_2^*) + \int_X R_{1,t+1}(y,\pi_1,\pi_2^*)Q(dy|x,a_1,f_2^*)\right]\pi_{1,t}(da_1|x,f_2^*).
\end{aligned}$$

Hence, by the induction hypothesis, it follows that

$$\begin{aligned}
R_{1,t}(x,\pi_1,\pi_2^*) &\le \int_{A_1}\left[P_{1,t}\beta_1^t r_1(x,a_1,f_2^*) + \int_X J_{1,t+1}(y)Q(dy|x,a_1,f_2^*)\right]\pi_{1,t}(da_1|x,f_2^*) \\
&\le \left(\int_{A_1}\pi_{1,t}(da_1|x,df_2^*)\right)\max_{a_1\in A_1(x)}\left[P_{1,t}\beta_1^t r_1(x,a_1,f_2^*) + \int_X J_{1,t+1}(y)Q(dy|x,a_1,f_2^*)\right] \\
&= \max_{a_1\in A_1(x)}\left[P_{1,t}\beta_1^t r_1(x,a_1,f_2^*) + \int_X J_{1,t+1}(y)Q(dy|x,a_1,f_2^*)\right] \\
&= J_{1,t}(x).
\end{aligned}$$

Therefore,

$$R_{1,t}(x,\pi_1,\pi_2^*) \le J_{1,t}(x), \quad \text{for all } x \in X.$$

This completes the inductive step.
2) On the other hand, if $R_{1,t+1}(x,\pi_1,\pi_2^*) = J_{1,t+1}(x), x \in X$ with $\pi_1 = \pi_1^*$, and $\pi_{1,t}(\cdot|h_t)$ is the Dirac measure concentrated at $f_{1,t}^*(x_t)$, then the equality in the previous calculations holds, yielding

$$R_{1,t}(x,\pi_1^*,\pi_2^*) = J_{i,t}(x)$$

In particular, since $R_{1,t}(x,\pi_1,\pi_2^*) \le J_{1,t}(x)$ for all t, we have at stage $t = 0$,

$$R_{1,0}(x,\pi_1,\pi_2^*) = \mathscr{J}_{1,\tau_1}(x,\pi_1,\pi_2^*) \le J_{1,0}(x)$$

and when $\pi_1 = \pi_1^*$,

$$\mathscr{J}_{1,\tau_1}(x,\pi_1^*,\pi_2^*) = J_{1,0}(x). \tag{4}$$

An analogous argument applies to player 2, assuming that player 1 acts rationally and follows the strategy π_1^*. Thus,

$$\mathscr{J}_{2,\tau_2}(x,\pi_1^*,\pi_2^*) = J_{2,0}(x). \tag{5}$$

Therefore, from equations (4) and (5), and since the maximum is attained, it coincides with the supremum. We conclude that

$$\mathscr{V}_{1,\tau_1}^*(x) = \sup_{\pi_1\in\Pi_1}\mathscr{J}_{1,\tau_1}(x,\pi_1,\pi_2^*), \quad \mathscr{V}_{2,\tau_2}^*(x) = \sup_{\pi_2\in\Pi_2}\mathscr{J}_{2,\tau_2}(x,\pi_1^*,\pi_2),$$

and hence $\pi^* = (\pi_1^*,\pi_2^*)$ constitutes a Nash equilibrium.

In conclusion, the construction of the functions $J_{i,t}$ and the existence of measurable selectors $f_{i,t}^* \in \mathbb{F}_i$ satisfying equation (3) establish part **a)**. The backward induction argument, combined with the verification of optimality through inequalities and equalities under fixed strategies π^*, confirms part **b)**, ensuring that $\pi^* = (\pi_1^*,\pi_2^*)$ constitutes a Nash equilibrium. □

Finally, we introduce a change of variables that simplifies the recursive structure of the dynamic programming equation. Let $U_{i,T_i+1}(x) := 0$ for all $x \in X$ and define

$$U_{i,t} := \frac{J_{i,t}}{P_{i,t}\beta_i^t}, t \in \{0,1,2,\dots,T_i-1\}.$$

Then the dynamic programming equation (3) is equivalent to

$$U_{i,t}(x) = \max_{a_i\in A_i(x)}\left[r_i(x,a_1,a_2) + \alpha_{i,t}\ \beta_i\int_X U_{i,t+1}(y)Q(dy|x,a_1,a_2)\right] \tag{6}$$

where

$$\alpha_{i,t} := \frac{P_{i,t+1}}{P_{i,t}}, t \in \{0,1,2,\dots,T_i - 1\}.$$

It is worth noting that the coefficient $\alpha_{i,t}$ admits a probabilistic interpretation: it corresponds to the conditional probability that the horizon of player i exceeds time $t+1$, given that it has not yet expired at time t. That is,

$$\alpha_{i,t} = \mathbb{P}(\tau_i \geq t+1 \mid \tau_i \geq t).$$

This factor modulates the contribution of future rewards in the dynamic programming recursion, accounting for the uncertainty in the effective duration of the game.

Moreover, in settings where the system dynamics are governed by a stochastic difference equation of the form

$$x_{t+1} = F(x_t, a_{1,t}, a_{2,t}, \xi_t), \quad t \geq 0,$$

with initial condition $x_0 = x \in X$, where $F : \mathbb{K} \times S \to X$ is a measurable function representing the system dynamics, and $\{\xi_t\}_{t\geq 0}$ is a sequence of independent and identically distributed random variables defined on a measurable space $(S, \mathscr{S})$. In this formulation, the transition kernel $Q(\cdot \mid x, a_1, a_2)$ is induced by the law of ξ_t through the mapping F, and the expectation E is taken with respect to the distribution of ξ_t. Under this representation, equation (6) can be equivalently written as

$$U_{i,t}(x) = \max_{a_i \in A_i(x)} \left[r_i(x, a_1, a_2) + \alpha_{i,t} \ \beta_i E\left[U_{i,t+1}(F(x, a_1, a_2, \xi_t))\right]\right], \tag{7}$$

for $t = 0, 1, ..., T_i$, taking into account that $U_{i,T_i+1}(x) = 0$.

The preceding analysis establishes a general framework for the existence and construction of equilibrium strategies in nonzero-sum Markov games with finite random horizons. In what follows, we illustrate the applicability of this framework through a concrete example concerning the strategic exploitation of a renewable resource. This case study, known as the Great Fish War game, has been widely used in the literature to model dynamic interactions between competing agents over a shared fishery resource.

4. The Great Fish War game with a Random Horizon

Consider a scenario in which two countries exploit fishing resources from a shared maritime region. Each country is characterized by its own utility function, which depends on the quantity of fish extracted in each period. Both players aim to maximize the expected sum of their discounted utilities over time, subject to a random horizon. Assuming that both countries act rationally, each takes into account the potential actions of the other when choosing its own strategy, in order to secure the most favorable extraction policy. Under this framework, consider that the state space is defined as $X = [0, \infty)$, representing the biomass of the shared fishery resource. The action sets for both players are given by $A_1 = A_2 = [0, \infty)$, corresponding to the feasible quantities of fish that can be extracted in a given period.

Moreover, at each state of the system, the natural growth of the biological resource is modeled by the fish population growth function $h : X \to [0, \infty)$, defined as $h(x) = x^\delta$, with $\delta > 0$. This function captures the regeneration capacity of the biomass as a function of the current population $x \in X$. Based on this growth law, the admissible actions sets are given by $A_1(x) = A_2(x) = \left[0, \frac{x^\delta}{M}\right]$, where $M \geq 1$, and $\frac{1}{M}$ represents the maximum extractable proportion per unit of biomass. This formulation implies that the maximum amount each country may harvest in a given period depends directly on the current state of the resource. In this way, the growth function not only governs the natural regeneration of the ecosystem, but also enforces a biological constraint on harvesting decisions, ensuring that exploitation remains within sustainable limits at all times.

Thus, the dynamics governing the evolution of the shared fishing resource between the two countries are described by the following difference equation:

$$x_{t+1} = (x_t - a_{1,t} - a_{2,t})^\delta \xi_t, t \geq 0,$$

where $x_0 = x \in X$ denotes the initial state of the system, and $a_{i,t} \in A_i(x_t)$ represents the quantity of fish extracted by player i in period $t \in \mathbb{N}_0$. The stock level at time t is denoted by x_t, and the condition $x_t > a_{1,t} + a_{2,t}$ is assumed to hold almost surely (a.s.) with respect to $\mathbb{P}_x^{\pi_1,\pi_2}$, ensuring that the joint extraction never exceeds the available biomass. This equation models the evolution of the fish population as a function of the extraction decisions made by both players. The term $h(x_t - a_{1,t} - a_{2,t}) = (x_t - a_{1,t} - a_{2,t})^\delta$ represents the surviving portion of the resource after joint extraction, while the multiplicative random factor ξ_t introduces a stochastic component that reflects environmental or biological disturbances, such as fluctuations in the natural mortality rate, diseases, migrations, or adverse conditions. The sequence $\{\xi_t\}$ is assumed to consist of independent and identically distributed random variables, with values in $S = [0, \infty)$, satisfying $E[\ln \xi_t] = \mu < \infty$, and with ξ_0 having a continuous density Δ.

Regarding the objectives of each player, individual reward functions are introduced to quantify the benefits obtained by each country in each period. For player i, the reward function $r_i : \mathbb{K} \to \mathbb{R} \cup \{-\infty, +\infty\}$ is given as

$$r_i(x, a_1, a_2) = \ln a_i, \quad \text{with } x \in X, \ a_i \in A_i(x).$$

This formulation indicates that the utility received by each player in a given period is proportional to the logarithm of the amount extracted. The use of a logarithmic function reflects the notion of diminishing marginal returns: as the quantity extracted increases, the additional utility obtained from each extra unit becomes smaller. This characteristic encourages more conservative and sustainable harvesting strategies in the long run.

The recursive structure of the Great Fish War game is depicted in Figure 1. At each stage, players observe the current state, select their actions, receive the corresponding payoffs, and then move to a new state determined by the transition law. This cycle repeats across successive stages,

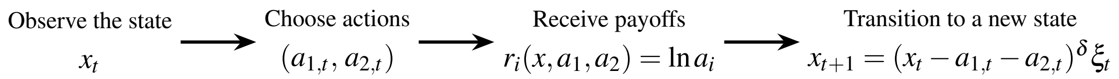

Observe the state
x_t
Choose actions
$(a_{1,t}, a_{2,t})$
Receive payoffs
$r_i(x, a_1, a_2) = \ln a_i$
Transition to a new state
$x_{t+1} = (x_t - a_{1,t} - a_{2,t})^{\delta}\xi_t$

Figure 1: Recursive dynamics of the Great Fish War game with two players.

giving rise to the recursive dynamics that characterize the game.

Additionally, each player i is assigned a random horizon τ_i, defined on the probability space $(\Omega_i, \mathscr{F}_i, \mathbb{P}_i)$. This random variable determines the duration over which player i participates in the game and models the uncertainty associated with the effective time available for resource exploitation. Since the state-action process evolves independently of the horizon variables, Assumption 1 is satisfied in this setting.

Now, we note that the Great Fish War game satisfies Assumption 2. Conditions (a) and (b) follow directly from the model construction. To verify condition (c), we now establish that the transition kernel $Q(\cdot \mid x, a_1, a_2)$ is weakly continuous in the actions. Fix $x \in X$, and for every measurable set $B \subset X$, the transition kernel is given by:

$$Q(B \mid x, a_1, a_2) = \int_0^{\infty} \mathbb{I}_B\left((x - a_1 - a_2)^{\delta} z\right) \Delta(z) dz.$$

Let $\psi_{x,a_1,a_2}(z) := (x - a_1 - a_2)^{\delta} z$ be the transformation of the random variable ξ. In accordance with the model assumptions, the condition $x > a_1 + a_2$ is satisfied at every stage of the game. Then, the function ψ_{x,a_1,a_2} is continuous in the action variables for any fixed state $x \in X$.

Let $(a_1^n, a_2^n) \to (a_1, a_2)$ be a sequence in $A(x)$. Then, for any bounded and continuous function $u : X \to \mathbb{R}$, we have:

$$\int_X u(y)\, Q(dy \mid x, a_1^n, a_2^n) = \int_0^{\infty} u\left((x - a_1^n - a_2^n)^{\delta} z\right) \Delta(z) dz.$$

Since $(a_1^n, a_2^n) \to (a_1, a_2)$, and u and Δ are continuous and bounded, the dominated convergence theorem implies:

$$\begin{aligned} \lim_{n\to\infty} \int_X u(y) Q(dy \mid x, a_1^n, a_2^n) &= \int_0^{\infty} u\left((x - a_1 - a_2)^{\delta} z\right) \Delta(z) dz \\ &= \int_X u(y) Q(dy \mid x, a_1, a_2). \end{aligned}$$

Hence, $Q(\cdot \mid x, a_1, a_2)$ is weakly continuous in (a_1, a_2), as required.

In this way, since Assumptions 1 and 2 are satisfied, the results of Theorem 1 can be applied. By solving the associated dynamic programming equations for this specific model, the functions of value and equilibrium strategies can be explicitly obtained. The results are summarized in the following proposition.

Proposition 1. The functions $U_{i,0}, U_{i,1}, \ldots, U_{i,T+1}$ for player i and the corresponding Nash equilibrium of the game with random horizon are given by

$$U_{1,t}(x) = K_t \ln x + C_t, \quad U_{2,t}(x) = \widehat{K}_t \ln x + \widehat{C}_t$$

and the equilibrium strategy profile is $\pi^* = (f_{1,t}^*(x), f_{2,t}^*(x))$ for each $x \in X$, where

$$f_{1,t}^*(x) = \frac{\widehat{K}_t - 1}{\widehat{K}_t K_t} x, \quad f_{2,t}^*(x) = \frac{K_t - 1}{\widehat{K}_t K_t} x,$$

and the coefficients satisfy the following backward recursions: $K_t = 1 + \alpha_{1,t}\beta_1\delta K_{t+1}$, $\widehat{K}_t = 1 + \alpha_{2,t}\beta_2\delta\widehat{K}_{t+1}$,

$$C_t = \ln\left(\frac{\widehat{K}_t - 1}{\widehat{K}_t K_t - 1}\right) + (K_t - 1)\ln\left(\frac{(\widehat{K}_t - 1)(K_t - 1)}{\widehat{K}_t K_t - 1}\right) + \alpha_{1,t}\beta_1[K_{t+1}\mu + C_{t+1}],$$

$$\widehat{C}_t = \ln\left(\frac{K_t - 1}{\widehat{K}_t K_t - 1}\right) + (\widehat{K}_t - 1)\ln\left(\frac{(\widehat{K}_t - 1)(K_t - 1)}{\widehat{K}_t K_t - 1}\right) + \alpha_{2,t}\beta_2[\widehat{K}_{t+1}\mu + \widehat{C}_{t+1}].$$

The explicit characterization obtained in Proposition 1 provides a solution to the Great Fish War game under a random horizon with finite support. The logarithmic structure of the value functions illustrates the diminishing marginal returns embedded in the utility specification, while the equilibrium strategies prescribe extraction policies proportional to the resource level. From a strategic perspective, the results show that the presence of uncertainty in the time horizon induces more conservative harvesting behavior, highlighting the role of environmental variability and temporal uncertainty in shaping optimal exploitation patterns.

5. Conclusion

The work focused on the formal development of the Markovian model involving rational players who make decisions under uncertainty regarding the duration of the game. The incorporation of a random horizon provided a more realistic framework for modeling situations in which the time span of the game is not predetermined, as often occurs in economic and ecological contexts.

An important outcome of this study was the formulation of a general framework for non-zero sum Markovian games with a random horizon. Under suitable structural assumptions, a dynamic programming approach was proposed to determine Nash equilibria. This method not only facilitated the proof of existence but also enabled a precise characterization of equilibrium strategies.

From an applied perspective, the Great Fish War game was analyzed under the framework of a random horizon with finite support. The explicit derivation of equilibrium strategies in this setting demonstrated how uncertainty about the duration of the interaction influences players' harvesting behavior. The results confirmed the capacity of the proposed framework to capture realistic dynamics in strategic resource exploitation under temporal uncertainty.

The proposed framework can also be applied to cost-based formulations. In this case, the objective shifts to minimizing cumulative costs, and the dynamic programming recursion involves infima instead of suprema. Accordingly, Assumption 2 must be adapted by requiring the stage cost functions to be lower semicontinuous. With these adjustments, the overall structure of the results remains valid. Several directions for further research remain open. One natural extension is to consider random horizons with infinite support, which pose additional technical challenges since the arguments used in the finite-support case are no longer applicable. Addressing this setting would likely require new analytical tools, such as weighted norms, to ensure convergence of the recursion and existence of equilibria. It would also be of interest to incorporate risk-sensitive criteria, incomplete information, or multi-player extensions, thereby enhancing the applicability of the model to more realistic scenarios.

References

[1] Bhabak, A., & Saha, S. (2022). Risk-sensitive semi-Markov decision problems with discounted cost and general utilities. *Statistics & Probability Letters, 184*, 109408. https://doi.org/10.1016/j.spl.2022.109408

[2] Camilo-Garay, C., Ortega-Gutiérrez, R. I., and Cruz-Suárez, H. (2020). "Optimal strategies for a fishery model applied to utility functions". *Mathematical Biosciences and Engineering*, 18(1):518–529. https://doi.org/10.3934/mbe.2021028

[3] Crespo-Guerrero, J. M. and Casado Izquierdo, J. M. (2023). "Pesca y comercialización del pulpo en Yucatán: ¿un proceso extractivista impulsado por la unión Europea?". *Geográfica Venezolana*, 64(2):301–319. https://doi.org/10.53766/RGV/2024.64.2.06

[4] Cruz-Suárez, H., Ilhuicatzi-Roldán, R., and Montes-de Oca, R. (2013). "Markov decision processes on borel spaces with total cost and random horizon". *Journal of Optimization Theory and Applications*, 162:329–346. https://doi.org/10.1007/s10957-012-0262-8

[5] Ekström, E. and Wang, Y. (2024). "Stopping problems with an unknown state". *Journal of Applied Probability*, 61(2):515–528. https://doi.org/10.1017/jpr.2023.52

[6] Gao, X., & Zhou, X. Y. (2024). Logarithmic regret bounds for continuous-time average-reward Markov decision processes. *SIAM Journal on Control and Optimization, 62*(5), 2529–2556. https://doi.org/10.48550/arXiv.2205.11168

[7] González-Sánchez, D., Luque-Vásquez, F., and Minjárez-Sosa, J. A. (2019). "Zero-sum markov games with random state-actions dependent discount factors: existence of optimal strategies". *Dynamic Games and Applications*, 9:103–121. https://doi.org/10.1007/s13235-018-0248-8

[8] Guo, X. and Wen, X. (2025). "Nonstationary nonzero-sum Markov games under a probability criterion". *arXiv preprint* arXiv:2505.10126. https://arxiv.org/abs/2505.10126

[9] Hernández-Lerma, O. and Lasserre, J. B. (1996). "Discrete-time Markov control processes: basic optimality criteria", volume 30. *Springer Science & Business Media*. https://doi.org/10.1007/978-1-4612-0729-0

[10] Jáskiewicz, A. and Nowak, A. S. (2016). "Non-zero-sum stochastic games". *In Handbook of dynamic game theory*, pages 281–344. Springer. https://doi.org/10.1007/978-3-319-27335-8_33-1

[11] Levhari, D. and Mirman, L. J. (1977). "Savings and consumption with an uncertain horizon". *Journal of Political Economy*, 85(2):265–281. https://doi.org/10.1086/260562

[12] Levhari, D. and Mirman, L. J. (1980). "The great fish war: an example using a dynamic Cournot-Nash solution". *The Bell Journal of Economics*, pages 322–334. https://doi.org/10.2307/3003416

[13] Miller, K. A. (2003). "North american pacific salmon: A case of fragile cooperation". *FAO Fisheries Report*, pages 105–122

[14] Minjárez-Sosa, J. A. (2020). "Zero-Sum discrete-time Markov games with unknown disturbance distribution: discounted and average criteria". *Springer Nature*. https://doi.org/10.1007/978-3-030-35720-7

[15] Missios, P. C. and Plourde, C. (1996). "The canada-european union turbot war: A brief game theoretic analysis". *Canadian Public Policy/Analyse de Politiques*, pages 144–150. https://doi.org/10.2307/3551905

[16] Nowak, A. (2006). "A note on an equilibrium in the great fish war game". *Economics Bulletin*, 17(2):1–10

[17] Ortega-Gutiérrez, R. I., Montes-de Oca, R., and Cruz-Suárez, H. (2024). "Characterization of a Cournot–Nash equilibrium for a fishery model with fuzzy utilities". *Journal of Mathematics*, 2024(1):6885051. https://doi.org/10.1155/2024/6885051

[18] Shapley, L. S. (1953). "Stochastic games". *Proceedings of the national academy of sciences*, 39(10):1095–1100. https://doi.org/10.1073/pnas.39.10.1095

[19] Wang, Y. (2023). "Optimal stopping, incomplete information, and stochastic games". Doctoral dissertation, Department of Mathematics, Uppsala University.